GENERALIZATION OF THE DERIVATIVE ANALYSIS METHOD TO NON-STATIONARY SINUSOIDAL MODELING

Sylvain Marchand
SCRIME / LaBRI – CNRS, Univ. Bordeaux 1
Talence, France
sylvain.marchand@labri.fr

Philippe Depalle
SPCL, Music Technology, McGill University
Montreal, Canada
depalle@music.mcgill.ca

ABSTRACT

In the context of non-stationary sinusoidal modeling, this paper introduces the generalization of the derivative method (presented at the first DAFx edition) for the analysis stage. This new method is then compared to the reassignment method for the estimation of all the parameters of the model (phase, amplitude, frequency, amplitude modulation, and frequency modulation), and to the Cramér-Rao bounds. It turns out that the new method is less biased, and thus outperforms the reassignment method in most cases for signal-to-noise ratios greater than -10dB.

1. INTRODUCTION

Sinusoidal sound modeling is widely used in many musical applications such as resynthesis, digital audio effects, transposition, time scaling [1], etc.

As regards the estimation of the model's parameters, many publications have focused on stationary sinusoidal analysis. In this context parameters of the partials are assumed not to evolve within an analysis frame. The comparison of the corresponding analysis methods is still an active research topic (see [2] for an example and references).

The extension of sinusoidal modeling to the non-stationary case improves quality when modeling the attacks or transients. While the synthesis is not a problem anymore (see [3]), the analysis in the presence of amplitude and/or frequency modulations remains a rather difficult task.

Non-stationary sinusoidal analysis has recently come back into light. Well-known analysis methods such as the quadratic interpolation [4, 5], the reassignment [6, 7], or the phase vocoder [8] have been generalized to the non-stationary case. Only the derivative method, which was proposed in the first edition of the DAFx conference by the first author, has not been generalized yet. This generalization is the aim of this paper.

After a brief presentation of non-stationary sinusoidal modeling in Section 2, we describe the reassignment method in Section 3 and we introduce the generalized derivative method in Section 4. We then compare these methods in theory (Section 5), in practice (Section 6), and against the Cramér-Rao lower bounds (CRBs).

2. NON-STATIONARY SINUSOIDAL MODELING

Additive synthesis can be considered as a spectrum modeling technique. It is originally rooted in Fourier's theorem, which states that any periodic function can be modeled as a sum of sinusoids at various amplitudes and harmonically related frequencies. In this paper we consider the sinusoidal model under its most general expression, which is a sum of complex exponentials (the *partials*) with time-varying amplitudes a_p and non-harmonically related frequencies ω_p (defined as the first derivative of the phases ϕ_p). The resulting signal s is thus given by:

$$s(t) = \sum_{p=1}^{P} a_p(t) \exp(\phi_p(t)). \tag{1}$$

In the context of this paper, amplitudes and frequencies are supposed to evolve within an analysis frame under first-order amplitude and frequency modulations. Furthermore, as the present study focuses on the statistical quality of the parameters' estimators rather than their frequency resolution, the signal model is reduced to only one partial ($P = 1$). The subscript notation for the partials is then useless. We also define Π_0 as being the value of the parameter Π at time 0, corresponding to the center of the analysis frame. The signal s is then given by:

$$s(t) = \exp\left(\underbrace{(\lambda_0 + \mu_0 t)}_{\lambda(t)=\log(a(t))} + j \underbrace{\left(\phi_0 + \omega_0 t + \frac{\psi_0}{2} t^2\right)}_{\phi(t)} \right) \tag{2}$$

where μ_0 (the amplitude modulation) is the derivative of λ (the log-amplitude), and ω_0 (the frequency), ψ_0 (the frequency modulation) are respectively, the first and second derivatives of ϕ (the phase). Thus, the log-amplitude and the phase are modeled by polynomials of degrees 1 and 2, respectively (see [3] for the corresponding synthesis method). These polynomial models can be viewed either as truncated Taylor expansions of more complicated amplitude and frequency modulations (*e.g.* tremolo / vibrato), or either as an extension of the stationary case where $\mu_0 = 0$ and $\psi_0 = 0$. Note that $a_0 = \exp(\lambda_0)$ and ϕ_0 are respectively the initial amplitude and initial phase of the signal.

3. NON-STATIONARY SINUSOIDAL ANALYSIS

The main problem we have to tackle now is the estimation of the model parameters, namely $a_0 = \exp(\lambda_0)$, μ_0, ϕ_0, ω_0, and ψ_0. This can be achieved, as in the stationary case, by using the short-time Fourier transform (STFT):

$$S_w(t,\omega) = \int_{-\infty}^{+\infty} s(\tau) w(\tau - t) \exp\left(-j\omega(\tau - t)\right) d\tau \tag{3}$$

where S_w is the short-time spectrum of the signal s. Note that we use here a slightly modified definition of the STFT than the usual one. Indeed we let the time reference slide with the window, which results in a phase shift of $-\omega t$. This is to simplify the comparison of the methods in the context of this paper; it does not however alter the quality of the estimations, all made at time $t = 0$. S_w involves an analysis window w, band-limited in such a way that for any frequency corresponding to one specific partial (corresponding to some local maximum in the magnitude spectrum), the influence of the other partials can be neglected (in the general case when $P > 1$).

In the stationary case ($\mu_0 = 0$ and $\psi_0 = 0$), the spectrum of the analysis window is simply centered on the frequency ω_0 and multiplied by the complex amplitude $s_0 = \exp(\lambda_0 + j\phi_0)$. In the non-stationary case however, s_0 gets multiplied [8] by:

$$\Gamma_w(\omega, \mu_0, \psi_0) = \int_{-\infty}^{+\infty} w(t) \exp\left(\mu_0 t + j\left(\omega t + \frac{\psi_0}{2} t^2\right)\right) dt. \tag{4}$$

Some preliminary attempts in non-stationary sinusoidal analysis consider the distortion of the resulting phase spectrum first from direct measurements [9], and then from polynomial approximations [10]. In the special case of using a Gaussian window for w, analytic formulas can be derived [11], and the quadratic interpolation method [4] is generalized by Abe and Smith [5]. The resulting so-called QIFFT method fits parabolas to the log-amplitude and phase spectra in order to estimate the model parameters. However, this method requires the unwrapping of the phase spectrum, zero padding, and is very sensitive to the shape of the window w (which in theory should be a Gaussian one – of infinite time support and with possibly a bad frequency resolution). As we do not want to impose the use of Gaussian windows nor zero padding in this present study, the QIFFT method might perform badly while another method, called the *reassignment*, seems to give better results (see [8]), at least as regards the estimation of the frequency.

The reassignment first proposed by Kodera, Gendrin, and de Villedary [12, 13], was generalized by Auger and Flandrin [6] for time and frequency. Hainsworth shows in [14] that the reassignment can be easily generalized for the amplitude modulation. Indeed, by considering Equation (3), one can easily derive:

$$\frac{\partial}{\partial t} \log\left(S_w(t, \omega)\right) = j\omega - \frac{S_{w'}(t, \omega)}{S_w(t, \omega)} \tag{5}$$

where w' denotes the derivative of w. From Equation (2), we then obtain the reassigned frequency $\hat{\omega}$ and amplitude modulation $\hat{\mu}$:

$$\hat{\omega}(t, \omega) = \frac{\partial}{\partial t} \Im\left(\log\left(S_w(t, \omega)\right)\right) = \omega - \underbrace{\Im\left(\frac{S_{w'}(t, \omega)}{S_w(t, \omega)}\right)}_{-\Delta_\omega}, \tag{6}$$

$$\hat{\mu}(t, \omega) = \frac{\partial}{\partial t} \Re\left(\log\left(S_w(t, \omega)\right)\right) = -\Re\left(\frac{S_{w'}(t, \omega)}{S_w(t, \omega)}\right). \tag{7}$$

In practice, for a partial p corresponding to a local maximum m of the (discrete) magnitude spectrum at the (discrete) frequency ω_m, the estimates of the frequency and the amplitude modulation are respectively given by:

$$\hat{\omega}_0 = \hat{\omega}(t, \omega_m) \quad \text{and} \quad \hat{\mu}_0 = \hat{\mu}(t, \omega_m). \tag{8}$$

Moreover, these instantaneous parameters are given for the reassigned time $\hat{t}_0 = \hat{t}(t, \omega_m)$, with:

$$\hat{t}(t, \omega) = t - \frac{\partial}{\partial \omega} \phi(t, \omega) = t + \underbrace{\Re\left(\frac{S_{tw}(t, \omega)}{S_w(t, \omega)}\right)}_{-\Delta_t}. \tag{9}$$

Normally these parameters might be reassigned to time $\hat{t}_0$. However, it appears in the experiments described in Section 6 that Δ_t can be neglected. Indeed, except for very low signal-to-noise ratios (SNRs) – in which case the estimation is not precise anyway – Δ_t is less than the sampling period (thus very low when compared to ω, μ, or ψ).

Nevertheless, Röbel has shown in [7] that, thanks to $\hat{t}$, an estimation of the frequency derivative ψ_0 is $\hat{\psi}_0 = \hat{\psi}(t, \omega_m)$ with:

$$\hat{\psi} = \frac{\partial \hat{\omega}}{\partial \hat{t}} = \frac{\partial \hat{\omega}}{\partial t} \Big/ \frac{\partial \hat{t}}{\partial t} \tag{10}$$

and

$$\frac{\partial \hat{\omega}}{\partial t} = \Im\left(\frac{S_{w''}}{S_w}\right) - \Im\left(\left(\frac{S_{w'}}{S_w}\right)^2\right), \tag{11}$$

$$\frac{\partial \hat{t}}{\partial t} = \Re\left(\frac{S_{tw} S_{w'}}{{S_w}^2}\right) - \Re\left(\frac{S_{tw'}}{S_w}\right). \tag{12}$$

Amplitude and phase are eventually to be estimated. Since these estimations are not included in the original reassignment method, and since we know the estimated modulations $\hat{\mu}_0$ and $\hat{\psi}_0$, we propose to use the Γ_w function of Equation (4), thus:

$$\hat{a}_0 = \left|\frac{S_w(\omega_m)}{\Gamma_w(\Delta_\omega, \hat{\mu}_0, \hat{\psi}_0)}\right|, \tag{13}$$

$$\hat{\phi}_0 = \angle\left(\frac{S_w(\omega_m)}{\Gamma_w(\Delta_\omega, \hat{\mu}_0, \hat{\psi}_0)}\right). \tag{14}$$

The reassignment method seems currently the best STFT-based method in terms of estimation precision, at least regarding frequency (see [8]).

4. GENERALIZED DERIVATIVE METHOD

At the first DAFx conference edition, the first author proposed in [15] to use the signal derivatives to estimate frequency and amplitude parameters of a sinusoidal model in the stationary case. We show here that the derivative method can be generalized to the non-stationary case.

4.1. Theoretical Considerations

More precisely, considering Equation (2), and since the derivative of an exponential is an exponential, we have:

$$s'(t) = (\mu_0 + j(\omega_0 + \psi_0 t)) \cdot s(t) \tag{15}$$

and thus

$$\Im\left(\frac{s'}{s}(t)\right) = \omega_0 + \psi_0 t \quad \text{and} \quad \Re\left(\frac{s'}{s}\right) = \mu_0. \tag{16}$$

For this method to work in the case of a signal made of several partials, we have to switch to the spectral domain. We consider only the spectrum values close to a given local magnitude maximum m (see Section 3) that represents the partial under investigation in order to be able to neglect the influence of the other partials.

We have to check first the contribution of $r(t) = \psi_0 t$ in the spectral domain. Fortunately, it has nice properties: Since $r(t)$ is an odd function, its spectrum $R(f)$ is imaginary, thus $j\psi_0 t$ only contributes to the real part of the spectrum of s'/s. Even-though its amplitude can be very high at extreme frequencies, $R(f)$ is null at frequency zero, and exhibits small values around 0.

Thus, the spectrum of s' involves a convolution sum between R, which equals 0 at frequency 0, with S_w, which energy is still essentially located around frequency ω_0 (it is exactly the case in the stationary case; it is only an approximation in the non-stationary case because of Equation (4) and the fact that the local maximum gets slightly shifted in frequency as shown by Abe and Smith in [5]). This convolution sum results in a negligible contribution when compared to $\omega_0 S_w$. The complete theoretical investigation of these properties is beyond the scope of this paper. However, in practice, evaluating S'_w/S_w close to the local maximum (discrete) frequency ω_m yields to an excellent estimation of the frequency:

$$\hat{\omega}_0 = \Im\left(\frac{S'_w}{S_w}(\omega_m)\right). \tag{17}$$

Now that we have gotten an estimate of ω_0, we can evaluate S'_w at this frequency, where $R * S_w$ contribution equals zero, and thus we obtain an estimate for the amplitude modulation:

$$\hat{\mu}_0 = \Re\left(\frac{S'_w}{S_w}(\hat{\omega}_0)\right). \tag{18}$$

In order to get the estimate of the frequency modulation ψ_0, we have to consider s'', the second derivative of s. More precisely, we have:

$$\begin{aligned} s''(t)/s(t) &= ({\mu_0}^2 - {\omega_0}^2 - 2\omega_0\psi_0 t - {\psi_0}^2 t^2) \\ &+ \; j(\psi_0 + 2\mu_0\omega_0 + 2\mu_0\psi_0 t). \end{aligned} \tag{19}$$

We then use the same kind of properties that we just used for the spectrum of the first derivative. Even functions (*e.g.* proportional to t^2) in one part (real or imaginary) of the signal will contribute to the same part (real or imaginary) of the spectrum, whether odd functions (*e.g.* proportional to t) in one part (real or imaginary) of the signal will contribute to the opposite part (imaginary or real) of the spectrum. Moreover the effects of the convolution sums are negligible for $\omega_m \approx \omega_0$. Finally, we get the estimate of the frequency modulation:

$$\hat{\psi}_0 = \Im\left(\frac{S''_w}{S_w}(\hat{\omega}_0)\right) - 2\hat{\mu}_0\hat{\omega}_0. \tag{20}$$

Let us consider now the estimation of the initial amplitude and initial phase of the signal. Since we know the estimated modulations $\hat{\mu}_0$ and $\hat{\psi}_0$, we propose to use the Γ_w function of Equation (4) as we did for the extension of the reassignment method in Section 3, but this time with $\Delta_\omega = \omega_0 - \hat{\omega}_0 \approx 0$, thus:

$$\hat{a}_0 = \left|\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right|, \tag{21}$$

$$\hat{\phi}_0 = \angle\left(\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right). \tag{22}$$

4.2. Practical Considerations

The last (but not least) problem is then in practice, to perform an estimation of the (discrete-time) derivatives s' and s'' from the (discrete-time) signal s. Unlike in [16], we will not reformulate the previous equations to adapt them to the discrete-time case, but instead, we will keep here with the mathematical definition:

$$s'(t) = \lim_{\epsilon \to 0} \frac{s(t+\epsilon) - s(t)}{\epsilon}. \tag{23}$$

Our first idea was to use a very accurate resampling method [17] to set ϵ very close to 0 in Equation (23). The two reconstructor filters (for $s(t+\epsilon)$ and $s(t)$) were then combined into a differentiator filter. But there are other ways to design differentiator filters. No matter the practical differentiator filter, as the order of this filter increases, its impulse response shall converge to the theoretical one. Indeed, differentiation is a linear operation, that can also be regarded as a filter of complex gain $j\omega$ (where ω is the frequency of the input sinusoid). The discrete-time response of this filter could be obtained by the inverse Fourier transform of its frequency response. However, we keep with the mathematical definitions. The continuous-time signal $s(t)$ can be reconstructed from its samples $s[m] = s(m/F_s)$ using the following equation:

$$s(t) = \sum_{m=-\infty}^{+\infty} s[m]\,\mathrm{sinc}(\underbrace{tF_s - m}_{u(t)}), \tag{24}$$

meaning that the signal $s'(t)$ is:

$$s'(t) = F_s \sum_{m=-\infty}^{+\infty} s[m]\left(\frac{\cos(\pi u(t))}{u(t)} - \frac{\sin(\pi u(t))}{\pi u(t)^2}\right) \tag{25}$$

which samples are (given that the multiplicative term of $s[m]$ in Equation (25) equals 0 when $m = n$):

$$s'[n] = s'\left(n/F_s\right) = F_s \sum_{m \neq n} s[m]\left(\frac{(-1)^{(n-m)}}{(n-m)}\right). \tag{26}$$

Thus, the discrete derivative s' can be obtained by convolving the discrete signal s by the following differentiator filter:

$$h[n] = F_s \frac{(-1)^n}{n} \text{ for } n \neq 0, \text{ and } h(0) = 0 \tag{27}$$

of infinite time support. In practice, we multiply h by the Hann window (see Equation (42)). This works quite well for high filter orders. With an order of 1023, the bias introduced by the approximation of the discrete derivative can be neglected. And the required convolution by an order-1023 impulse response is quite fast on nowadays computers. Figure 1 shows the maximal relative

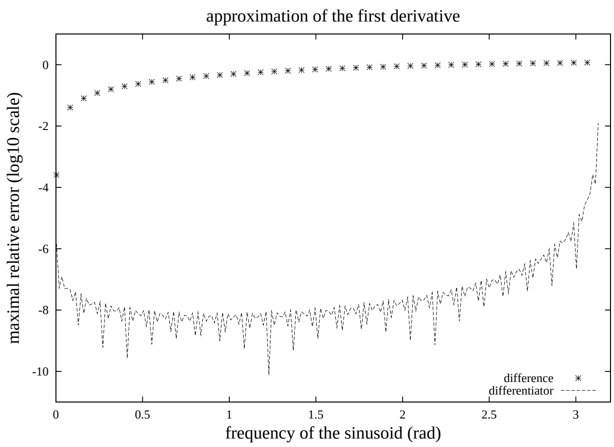


Figure 1: *Maximal estimation error when approximating the (first) derivative of a sinusoid using the order-1023 differentiator. The error is acceptable for all but very high frequencies. As a comparison, the error obtained with the classic difference approximation (equivalent to choosing $\epsilon = 1/F_s$ in the expression of Equation (23)) is plotted with stars $*$.*

error for the first derivative, when the signal is a pure sinusoid of constant (normalized) frequency $\underline{\omega}$ and amplitude 1. The relative error is the maximal absolute error divided by the norm of the gain of the theoretic derivative ($\underline{\omega}$). These results show that the errors remain acceptable for most frequencies, except very high frequencies close to the (normalized) Nyquist frequency (π). To obtain the second derivative, the differentiation is applied twice.

Finally, for each analysis frame, whereas the reassignment method requires 5 fast Fourier transforms (FFTs), the derivative method requires only 3 of them. When the derivatives are to be estimated using the differentiator, an efficient implementation using a larger FFT size is still possible (but out of the scope of this paper, which focuses on the precision of the estimations).

5. THEORETICAL EQUIVALENCES

In [2], the reassignment (Section 3) and derivative (Section 4) methods are proven to be theoretically equivalent, at least as regards the estimation of the frequency in the stationary case.

In order to prove similar equivalences in the non-stationary case, we apply the same strategy than in [2] and thus we introduce $\rho = \tau - t$ which gives another (equivalent) expression for the STFT (see Equation (3)):

$$S_w(t,\omega) = \int_{-\infty}^{+\infty} s(t+\rho)w(\rho)\exp\left(-j\omega\rho\right)\, d\rho \tag{28}$$

from which we can derive:

$$\frac{\partial}{\partial t}\log\left(S_w(t,\omega)\right) = \frac{S'_w(t,\omega)}{S_w(t,\omega)}. \tag{29}$$

By considering Equation (29) instead of Equation (5) in Section 3, we would have obtained:

$$\hat{\omega} = \Im\left(\frac{S'_w}{S_w}\right), \tag{30}$$

$$\hat{\mu} = \Re\left(\frac{S'_w}{S_w}\right) \tag{31}$$

instead of Equations (6) and (7). Finally, when considering Equations (17) and (18), we can conclude that the reassignment and derivative methods are equivalent, at least in theory and for the estimation of the frequency and the amplitude modulation. However, Section 6 will show differences in practice, since their practical implementations are very different.

6. PRACTICAL EXPERIMENTS AND RESULTS

In this section, we quantitatively evaluate the precision of the reassignment and derivative methods for the estimation of all the model parameters, and compare them to theoretical lower bounds.

For these experiments, we consider discrete-time signals s, with sampling rate F_s, consisting of 1 complex exponential generated according to Equation (2) with an initial amplitude $a_0 = 1$, and mixed with a Gaussian white noise of variance σ^2.

The analysis frames we consider are of odd length $N = 2H + 1$ samples, with the estimation time 0 set at their center.

In Equations (3) and (4), the continuous integrals turn into discrete summations over N values, with indices from $-H$ to $+H$.

In this section, the parameters are normalized to make them independent of the sampling frequency F_s. We thus define $\underline{\mu} = \mu/F_s$, $\underline{\omega} = \omega/F_s$, and $\underline{\psi} = \psi/{F_s}^2$.

6.1. Theoretical Bounds

When evaluating the performance of an estimator in the presence of noise and in terms of the variance of the estimation error, an interesting element to compare with is the Cramér-Rao bound (CRB). The CRB is defined as the limit to the best possible performance achievable by an unbiased estimator given a data set. For the model of Equation (2), for the five model parameters, these bounds have been derived by Zhou *et al.* [18]. We will consider their asymptotic versions (for a large N and a high number of observations).

Djurić and Kay [19] have shown that the CRBs depend on the time sample n_0 at which the parameters are estimated, and that the optimal choice in terms of lower bounds is to set n_0 at the center of the frame, *i.e.* $n_0 = H$, since the CRBs depend on:

$$\epsilon_k(\underline{\mu}, N) = \sum_{n=0}^{N-1}\left(\frac{n-n_0}{N}\right)^k \exp\left(2\underline{\mu}\frac{n-n_0}{N}\right). \tag{32}$$

6.1.1. Amplitude and Amplitude Modulation

After Zhou *et al.* [18], we define:

$$D_1(\underline{\mu}, N) = 2(\epsilon_0\epsilon_2 - {\epsilon_1}^2) \tag{33}$$

and give the expressions of the bounds for the amplitude a and amplitude modulation $\underline{\mu}$:

$$\mathrm{CRB}_{a,N}(\sigma,\underline{\mu}) \approx \frac{\sigma^2\epsilon_2}{D_1}, \tag{34}$$

$$\mathrm{CRB}_{\underline{\mu},N}(\sigma,a,\underline{\mu}) \approx \frac{\sigma^2\epsilon_0}{a^2N^2D_1}. \tag{35}$$

6.1.2. Phase, Frequency, and Frequency Modulation

As explained by Zhou *et al.* [18], the expressions of the bounds are different whether there is a frequency modulation or not (because this changes the degree of the polynomial associated to the phase).

In the absence of frequency modulation ($\psi = 0$), the bounds for the phase ϕ and frequency $\underline{\omega}$ are given by:

$$\mathrm{CRB}_{\phi,N}(\sigma,a,\underline{\mu}) \approx \frac{\sigma^2\epsilon_2}{a^2D_1}, \tag{36}$$

$$\mathrm{CRB}_{\underline{\omega},N}(\sigma,a,\underline{\mu}) \approx \frac{\sigma^2\epsilon_0}{a^2N^2D_1}. \tag{37}$$

In the presence of frequency modulation, the expressions of the bounds for the phase (ϕ), frequency ($\underline{\omega}$), and frequency modulation ($\underline{\psi}$) are given by:

$$\mathrm{CRB}_{\phi,N}(\sigma,a,\underline{\mu}) \approx \frac{\sigma^2(\epsilon_2\epsilon_4 - {\epsilon_3}^2)}{a^2D_2}, \tag{38}$$

$$\mathrm{CRB}_{\underline{\omega},N}(\sigma,a,\underline{\mu}) \approx \frac{\sigma^2(\epsilon_0\epsilon_4 - {\epsilon_2}^2)}{a^2N^2D_2}, \tag{39}$$

$$\mathrm{CRB}_{\underline{\psi},N}(\sigma,a,\underline{\mu}) \approx 4\frac{\sigma^2(\epsilon_0\epsilon_2 - {\epsilon_1}^2)}{a^2N^4D_2}, \tag{40}$$

where $D_2(\underline{\mu}, N)$ is defined as:

$$D_2(\underline{\mu}, N) = 2(\epsilon_0\epsilon_2\epsilon_4 - {\epsilon_1}^2\epsilon_4 - \epsilon_0{\epsilon_3}^2 + 2\epsilon_1\epsilon_2\epsilon_3 - {\epsilon_2}^3). \tag{41}$$

6.2. Practical Experiments

In our experiments, we set $F_s = 44100\text{Hz}$, $N = 511$, and the signal-to-noise ratio (SNR) expressed in dB is $10\log_{10}(\sigma^2/a^2)$, and goes from -20dB to $+100\text{dB}$ by steps of 5dB.

For each SNR and for each analysis method, we test 99 frequencies (ω) linearly distributed in the $(0, 3F_s/8)$ interval, and 9 phases (ϕ) linearly distributed in the $(-\pi, +\pi)$ interval. The amplitude modulation (μ) is either 0 (no-AM case) or one of 5 values linearly distributed in the $[-100, +100]$ interval (AM case). The frequency modulation (ψ) is either 0 (no-FM case) or one of 5 values linearly distributed in the $[-10000, +10000]$ interval (FM case). We limit the frequency to $3/4$ of the Nyquist frequency because very high frequencies are problematic (cause bias) for the derivative method (see Section 4). This is not really problematic in practice, since with a sampling rate of 44100Hz, this mostly covers the range of audible frequencies. Moreover, nowadays sampling frequencies can be as high as 96kHz, or even 192kHz.

For the analysis window w, we use the zero-centered (symmetric) Hann window of duration $T = N/F_s$ – that is of (odd) size N samples, defined for continuous time by:

$$w(t) = \frac{1}{2}\left(1 + \cos\left(2\pi\frac{t}{T}\right)\right). \tag{42}$$

We then compare the reassignment method (R) – see Section 3 – and two variants of the derivative method (D) – see Section 4: The estimated derivative method (ED), where the derivatives s' and s'' are estimated using the differentiator filter described in Section 4; and the theoretic derivative method (TD), where the exact derivatives s' and s'' are given by Equations (15) and (19). As regards the noise, the derivatives are approximated by using the differentiator filter. We consider the TD method because the estimated derivatives can be improved (*e.g.* by increasing the order of the differentiator filter). Thus the results of the TD method can be regarded as the best performance the D method could achieve, though with a better approximation of the discrete derivatives.

As mentioned previously, the comparison with QIFFT [5] is left apart in this paper and part of our future work. However, the comparison of reassignment with QIFFT [5] in various situations can be found in [8], but only for the estimation of the frequency.

It is also important to note that Badeau *et al.* have shown in [20] that the high-resolution ESPRIT method was close to the CRBs. But this method is very sensitive to the choice of the model order and do not consider frequency modulation ($\psi = 0$).

6.3. Experimental Results

When looking at the results of these experiments (see Figures 2–6, it turns out that with the proposed differentiator filter, the order of 1023 is sufficient for the method to achieve in practice (ED) performances very close to the theory (TD).

Below -10dB (high-error range), the derivative method appears to make larger errors than the reassignment method (except for phase and frequency, where the two methods behave similarly). However, this area is of poor practical interest since the errors are too important.

Above -10dB, as the SNR increases the reassignment method gets almost always biased, whereas the derivative method achieves performances closer to the CRB in nearly all cases. When FM is present, the derivative method is better for the estimation of the amplitude (see Figure 2), amplitude modulation (see Figure 3), and phase (see Figure 4); both methods are equivalent for the estimation of the frequency (see Figure 5); and the reassignment method is better for the estimation of the frequency modulation (see Figure 6), while the error of the derivative method is very low.

Moreover, we have observed the following trends: For small frame sizes, the derivative method outperforms the reassignment method in all cases. When the frame size increases, both methods are quite equivalent without FM, even if the order of the differentiator might have to be increased to lower the bias for the estimation of the amplitude modulation and frequency.

7. CONCLUSION AND FUTURE WORK

In this paper, we have generalized the derivative method for non-stationary sinusoidal modeling, and compared it to the reassignment method, which is currently among the best analysis methods. We have shown that, in theory, these two methods are equivalent for the estimation of the frequency and amplitude modulation. In practice, for signal-to-noise ratios above -10dB and without considering very high frequencies, the derivative method outperforms the reassignment method in all cases except for the estimation of the frequency modulation.

8. ACKNOWLEDGMENTS

Sylvain Marchand would like to thank Prof. Thomas F. Quatieri for the great discussion about non-stationary sinusoidal modeling after the talk for [3] at IEEE WASPAA 2003. Although we were then already convinced that the derivative method could work in the non-stationary case, it took some years to find some time...

Many thanks also go to Roland Badeau and Prof. Julius O. Smith III for the fruitful discussions about differentiator filters.

This research was partly supported by the French GIP ANR, DESAM project (ANR-06-JCJC-0027-01).

9. REFERENCES

[1] Sylvain Marchand and Martin Raspaud, “Enhanced Time-Stretching Using Order-2 Sinusoidal Modeling,” in *Proceedings of the Digital Audio Effects (DAFx) Conference*, Naples, Italy, October 2004, pp. 76–82.

[2] Mathieu Lagrange and Sylvain Marchand, “Estimating the Instantaneous Frequency of Sinusoidal Components Using Phase-Based Methods,” *Journal of the Audio Engineering Society*, vol. 55, no. 5, pp. 385–399, May 2007.

[3] Laurent Girin, Sylvain Marchand, Joseph di Martino, Axel Röbel, and Geoffroy Peeters, “Comparing the Order of a Polynomial Phase Model for the Synthesis of Quasi-Harmonic Audio Signals,” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, October 2003, pp. 193–196.

[4] Julius O. Smith III and Xavier Serra, “PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds based on a Sinusoidal Representation,” in *Proceedings of the International Computer Music Conference (ICMC)*, Champaign-Urbana, USA, 1987, pp. 290–297.

[5] Mototsugu Abe and Julius O. Smith III, “AM/FM Rate Estimation for Time-Varying Sinusoidal Modeling,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, NY, USA, March 2005, vol. III, pp. 201–204.

[6] François Auger and Patrick Flandrin, “Improving the Readibility of Time-Frequency and Time-Scale Representations by the Reassignment Method,” *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[7] Axel Röbel, "Estimating Partial Frequency and Frequency Slope Using Reassignment Operators," in *Proceedings of the International Computer Music Conference (ICMC)*, Göteborg, Sweden, September 2002, pp. 122–125.

[8] Michaël Betser, Patrice Collen, Gaël Richard, and Bertrand David, "Estimation of Frequency for AM/FM Models Using the Phase Vocoder Framework," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 505–517, February 2008.

[9] Paul Masri and Nishan Canagarajah, "Extracting More Detail From the Spectrum With Phase Distortion Analysis," in *Proceedings of the Digital Audio Effects (DAFx) Workshop*, Barcelona, Spain, November 1998, pp. 119–122.

[10] Mathieu Lagrange, Sylvain Marchand, and Jean-Bernard Rault, "Sinusoidal Parameter Extraction and Component Selection in a Non Stationary Model," in *Proceedings of the Digital Audio Effects (DAFx) Conference*, Hamburg, Germany, September 2002, pp. 59–64.

[11] Jorge S. Marques and Luis B. Almeida, "A Background for Sinusoid Based Representation of Voiced Speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Tokyo, Japan, April 1986, vol. 11, pp. 1233–1236.

[12] K. Kodera, C. de Villedary, and R. Gendrin, "A New Method for the Numerical Analysis of Time-Varying Signals with Small BT Values," *Phys. Earth Planet. Interiors*, , no. 12, pp. 142–150, 1976.

[13] K. Kodera, R. Gendrin, and C. de Villedary, "Analysis of Time-Varying Signals with Small BT Values," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 1, pp. 64–76, 1986.

[14] Stephen W. Hainsworth, *Techniques for the Automated Analysis of Musical Audio*, Ph.D. thesis, University of Cambridge, United Kingdom, December 2003.

[15] Sylvain Marchand, "Improving Spectral Analysis Precision With an Enhanced Phase Vocoder Using Signal Derivatives," in *Proceedings of the Digital Audio Effects (DAFx) Workshop*, Barcelona, Spain, November 1998, pp. 114–118.

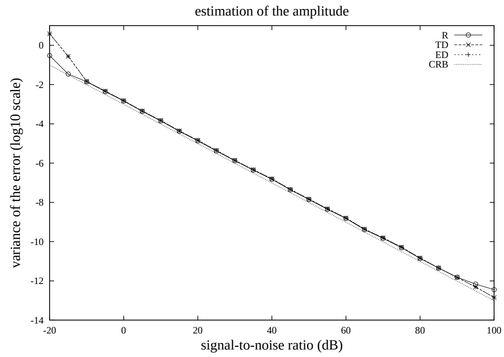
[16] Myriam Desainte-Catherine and Sylvain Marchand, "High Precision Fourier Analysis of Sounds Using Signal Derivatives," *Journal of the Audio Engineering Society*, vol. 48, no. 7/8, pp. 654–667, July/August 2000.

[17] Martin Raspaud and Sylvain Marchand, "Enhanced Resampling for Sinusoidal Modeling Parameters," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, October 2007.

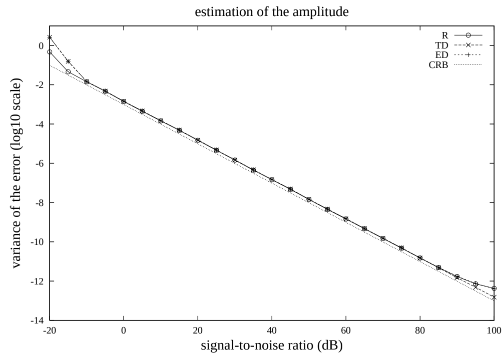
[18] Guotong Zhou, Georgios B. Giannakis, and Ananthram Swami, "On Polynomial Phase Signal with Time-Varying Amplitudes," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 848–860, April 1996.

[19] Petar M. Djurić and Steven M. Kay, "Parameter Estimation of Chirp Signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2118–2126, December 1990.

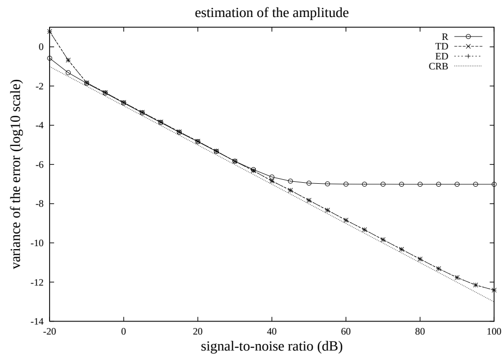
[20] Roland Badeau, Gaël Richard, and Betrand David, "Performance of ESPRIT for Estimating Mixtures of Complex Exponentials Modulated by Polynomials," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 492–504, February 2008.



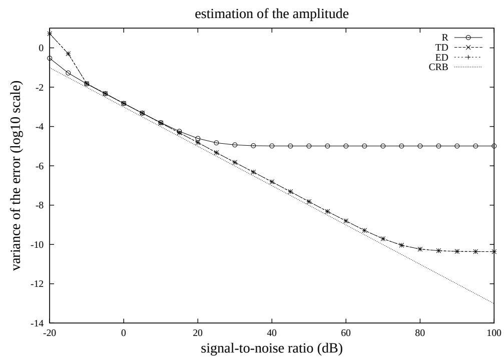
(a) stationary case ($\mu = 0$,$\psi = 0$)



(b) AM-only case ($|\mu| \leq 100$,$\psi = 0$)

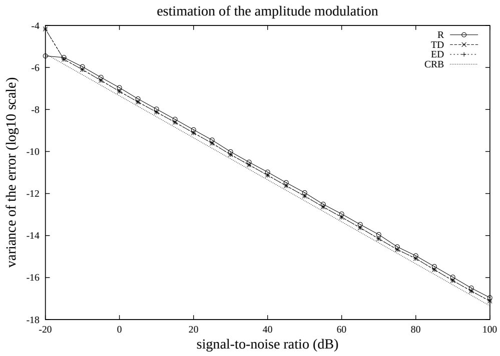


(c) FM-only case ($\mu = 0$,$|\psi| \leq 10000$)

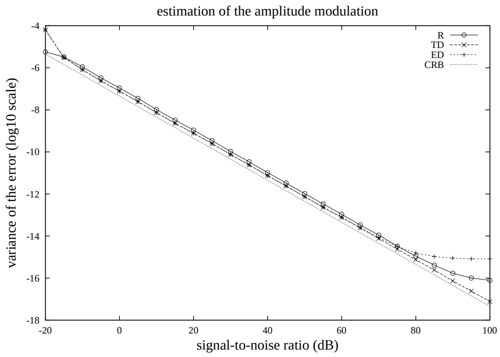


(d) AM/FM case ($|\mu| \leq 100$,$|\psi| \leq 10000$)

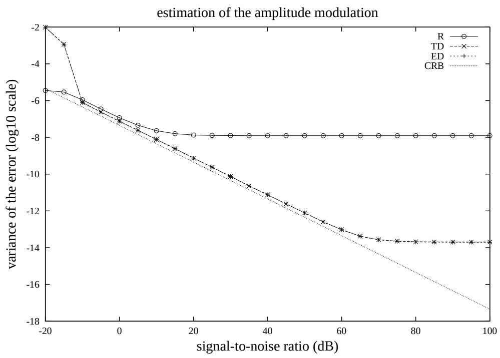
Figure 2: *Amplitude estimation error as a function of the SNR (stationary, AM-only, FM-only, and AM/FM cases) for the R, ED, and TD methods, and comparison to CRB$_a$.*



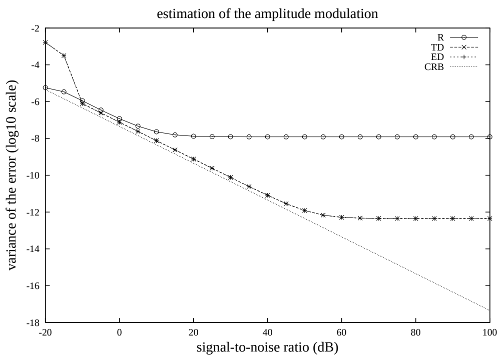
(a) stationary case ($\mu = 0$,$\psi = 0$)



(b) AM-only case ($|\mu| \leq 100$,$\psi = 0$)

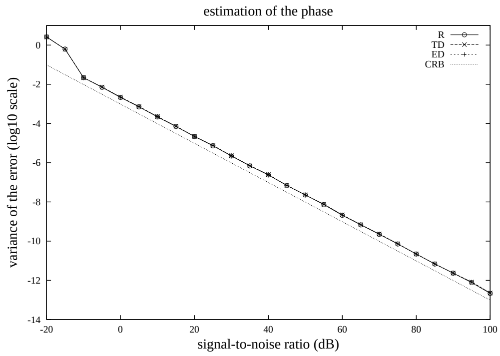


(c) FM-only case ($\mu = 0$,$|\psi| \leq 10000$)

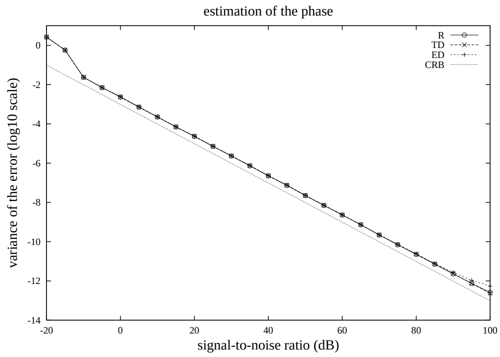


(d) AM/FM case ($|\mu| \leq 100$,$|\psi| \leq 10000$)

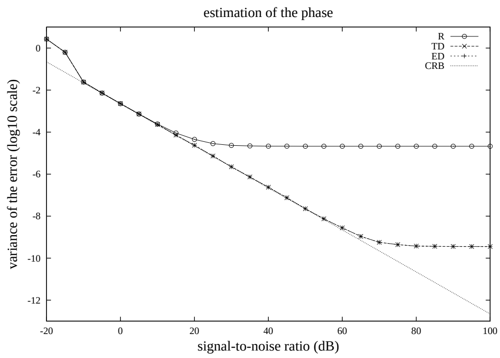
Figure 3: *Amplitude modulation estimation error as a function of the SNR (stationary, AM-only, FM-only, and AM/FM cases) for the R, ED, and TD methods, and comparison to CRB$_{\underline{\mu}}$.*



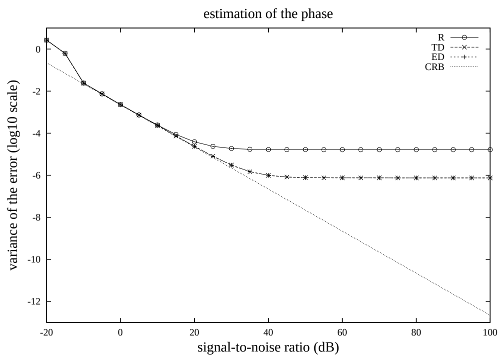
(a) stationary case ($\mu = 0$,$\psi = 0$)



(b) AM-only case ($|\mu| \leq 100$,$\psi = 0$)

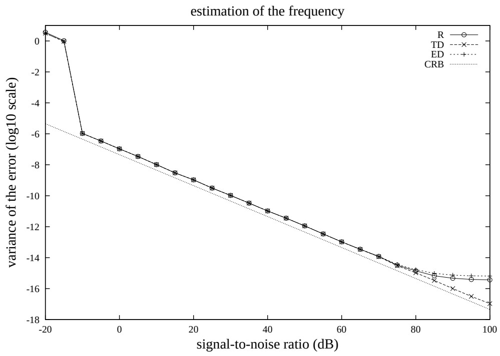


(c) FM-only case ($\mu = 0$,$|\psi| \leq 10000$)

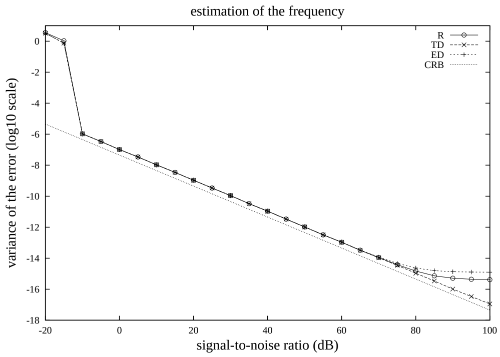


(d) AM/FM case ($|\mu| \leq 100$,$|\psi| \leq 10000$)

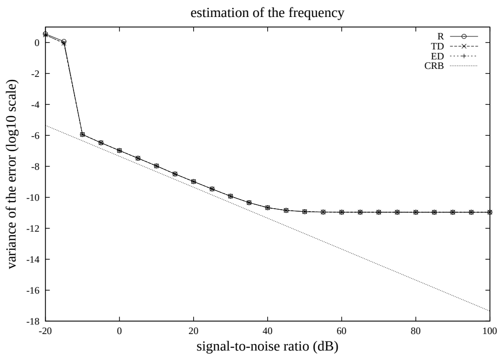
Figure 4: *Phase estimation error as a function of the SNR (stationary, AM-only, FM-only, and AM/FM cases) for the R, ED, and TD methods, and comparison to CRB$_\phi$.*



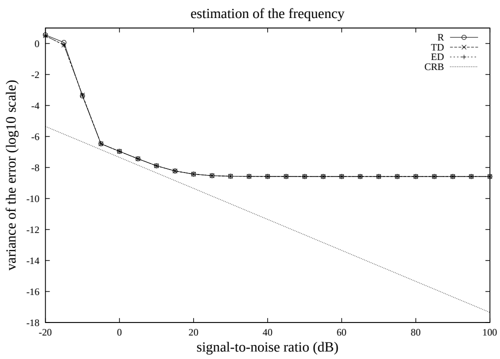
(a) stationary case ($\mu = 0$,$\psi = 0$)



(b) AM-only case ($|\mu| \leq 100$,$\psi = 0$)

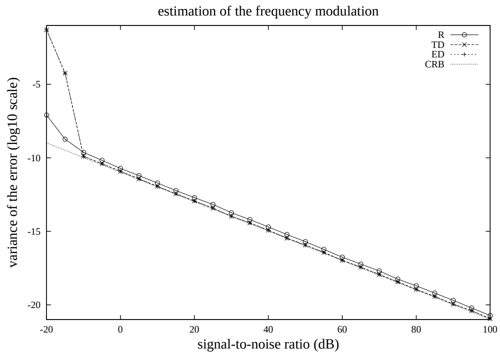


(c) FM-only case ($\mu = 0$,$|\psi| \leq 10000$)

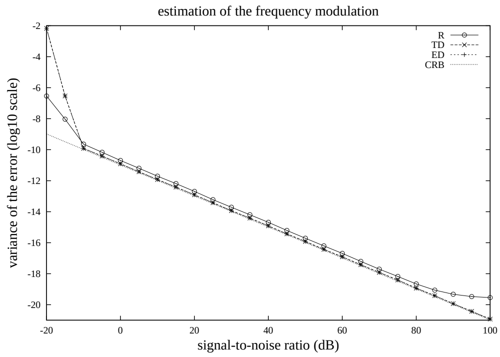


(d) AM/FM case ($|\mu| \leq 100$,$|\psi| \leq 10000$)

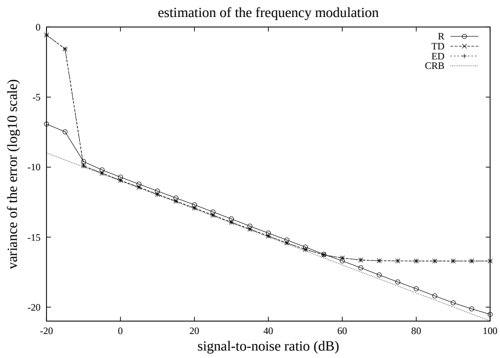
Figure 5: *Frequency estimation error as a function of the SNR (stationary, AM-only, FM-only, and AM/FM cases) for the R, ED, and TD methods, and comparison to CRB*$_{\underline{\omega}}$.



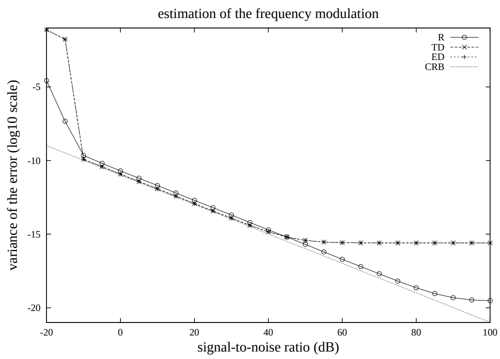
(a) stationary case ($\mu = 0$,$\psi = 0$)



(b) AM-only case ($|\mu| \leq 100$,$\psi = 0$)



(c) FM-only case ($\mu = 0$,$|\psi| \leq 10000$)



(d) AM/FM case ($|\mu| \leq 100$,$|\psi| \leq 10000$)

Figure 6: *Frequency modulation estimation error as a function of the SNR (stationary, AM-only, FM-only, and AM/FM cases) for the R, ED, and TD methods, and comparison to CRB*$_{\underline{\psi}}$.