THE SIMPLEST ANALYSIS METHOD FOR NON-STATIONARY SINUSOIDAL MODELING

Sylvain Marchand
Lab-STICC – CNRS, University of Brest, France
`Sylvain.Marchand@univ-brest.fr`

ABSTRACT

This paper introduces an analysis method based on the generalization of the phase vocoder approach to non-stationary sinusoidal modeling. 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 this method compares to the state of the art in terms of performances, with the great advantage of being much simpler.

1. INTRODUCTION

Sinusoidal modeling can be successfully used in digital audio effects such as pitch shifting or time scaling. It has recently gained a new interest with its extension to the non-stationary case, where amplitude / frequency modulations are taken into consideration.

The quality of the sound depends mainly on the analysis stage, where the model parameters are extracted from real sounds. At the first DAFx edition, I proposed to use the derivatives of the signal [1]. This analysis method was later generalized to the non-stationary case [2, 3]. Non-stationary sinusoidal analysis has recently become very active, with the works of Betser [4, 5, 6], Hamilton and Depalle [7, 8, 9], Wen and Sandler [10], or Muševič and Bonada [11, 12]. This has brought efficient methods, and a deep understanding of how these complex methods are related.

However, to gain wide acceptance, a method has to be not only efficient, but also simple and robust. Some methods (*e.g.* high-resolution methods) are very precise but may perform poorly if the signal model is not perfectly respected. Other methods lack accuracy (*e.g.* simple peak picking in the Fourier spectrum), but are widespread because of their simplicity…

In this paper, I propose to generalize the well-known phase vocoder approach to the non-stationary case, in a very simple way. This work must be seen as an extension of [2], with a new – simple – method but the same aims and experiments. The estimation is based on the (discrete) Fourier spectrum and uses derivatives approximated by first-order differences. This makes the method robust and fast (Fourier spectrum, FFT-based), simple (phase vocoder approach, first-order differences), and yet quite efficient (in comparison to state-of-the-art techniques or theoretical bounds).

After a presentation of non-stationary sinusoidal modeling in Section 2, Section 3 makes a short survey on derivative-based analysis. The simple generalized difference method is introduced in Section 4. This method is then compared to the reassignment method and against the Cramér-Rao lower bounds in Section 5.

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. Let us consider here 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:

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

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. Let us 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:

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

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, which 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 $\psi_0 = 0$.

3. SINUSOIDAL ANALYSIS

The problem we are interested in 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 using the short-time Fourier transform (STFT):

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

where S_w is the short-time spectrum of the signal s. Note that, as in [2], we use here a slightly modified definition of the STFT. Indeed we let the time reference slide with the window, which is also the case in practice when the STFT is implemented using a sliding FFT. 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 m 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

$$s_0 = a_0 \exp(j\phi_0) = \exp(\lambda_0 + j\phi_0). \quad (4)$$

In the non-stationary case however, s_0 gets multiplied by

$$\Gamma_w(\omega, \mu_0, \psi_0) = \int_{-\infty}^{+\infty} w(t) \exp\left(\mu_0 t + j\left(\omega t + \frac{\psi_0}{2} t^2\right)\right) dt. \quad (5)$$

In the special case of using a Gaussian window for w, an analytic formula can be derived [13]. This is also feasible, but more complicated, for other windows [11]. Else, it is always possible – although time consuming – to compute Γ_w directly from Equation (5). Once the estimated frequency $\hat{\omega}_0$ and modulations $\hat{\mu}_0$, $\hat{\psi}_0$ are known, the amplitude and phase can eventually be estimated since

$$\hat{s}_0 = \frac{S_w(t, \omega_m)}{\Gamma_w(\hat{\omega}_0 - \omega_m, \hat{\mu}_0, \hat{\psi}_0)}, \quad (6)$$

where ω_m is the (discrete) frequency of the local maximum of the (discrete) magnitude spectrum where the partial is detected.

The problem is yet to estimate the frequency and the modulations. The frequency modulation is the derivative of the frequency. And the amplitude modulation and frequency can be estimated using the derivatives of the real and imaginary parts, respectively, of the logarithm of the spectrum (see [2]). Indeed, these real and imaginary parts of the spectrum are the log-amplitude and phase, respectively, since in the (log-)polar notation $S_w = a\exp(j\phi) = \exp(\lambda + j\phi)$. This estimation can involve either the derivatives of the signal (derivative method) or the derivatives of the analysis window (reassignment method). The equivalence of these two approaches can be shown by a change of variable in the integral of Equation (3), see [2], or by integration by parts, see [10].

The last problem is now to be able to compute the derivative of the signal or of the window. Sometimes, this can be done analytically. Else, it is possible to use a differentiator filter [2]. But a much simpler way is to approximate the derivative using the (first-order) difference. This is the case in the phase vocoder, where the instantaneous frequency is estimated from the phase difference.

4. THE SIMPLEST METHOD

Let us generalize the phase vocoder approach to the non-stationary case, considering frames x of N consecutive samples of the signal s, and their discrete spectra X obtained by zero-phase Discrete Fourier Transform (DFT). Thus $X(\omega) = S_w(t, \omega)$ is the spectrum of the frame centered at the desired (discrete) estimation time, and let $X_{\mp}(\omega) = S_w(t \mp 1/F_s, \omega)$ be its left (previous, *i.e.* one sample before) and right (next, *i.e.* one sample after) neighboring spectra, respectively (F_s denoting the sampling frequency). Thus, the derivative is approximated by the first-order difference. As used in [14], let us notice that the log-amplitude and phase differences correspond to the real and imaginary parts of the logarithm of spectral ratios, respectively, and define:

$$\begin{aligned} \Delta_\lambda(X_1, X_2) &= \log|X_1| - \log|X_2| \\ &= \Re\left(\log\left(X_1/X_2\right)\right), \quad (7) \\ \Delta_\phi(X_1, X_2) &= \angle X_1 - \angle X_2 \\ &= \Im\left(\log\left(X_1/X_2\right)\right) \quad (8) \end{aligned}$$

(X_1 and X_2 denoting two complex spectra).

Since we can measure the amplitude of the spectra, we can compute the left and right estimates of the amplitude modulation, and retain their mean as the final estimation:

$$\begin{aligned} \mu_- &= \Delta_\lambda(X, X_-) \cdot F_s, \quad (9) \\ \mu_+ &= \Delta_\lambda(X_+, X) \cdot F_s, \quad (10) \\ \hat{\mu}_0 &= (\mu_- + \mu_+)/2. \quad (11) \end{aligned}$$

Similarly, with the measured phase of the spectra, we can compute an estimation of the instantaneous frequency:

$$\begin{aligned} \omega_- &= \text{unwrap}\left(\Delta_\phi(X, X_-)\right) \cdot F_s, \quad (12) \\ \omega_+ &= \text{unwrap}\left(\Delta_\phi(X_+, X)\right) \cdot F_s, \quad (13) \\ \hat{\omega}_0 &= (\omega_- + \omega_+)/2, \quad (14) \end{aligned}$$

where unwrap(α) is the function consisting in adding 2π to α if it is lower than 0 (*i.e.* the same phase unwrapping procedure as in the phase vocoder). With the left and right estimates of the frequency, again by first-order difference we can derive an estimate of the frequency modulation:

$$\hat{\psi}_0 = (\omega_+ - \omega_-) \cdot F_s. \quad (15)$$

Finally, to get the amplitude and phase estimates, we apply the procedure described in the previous section (see Equations (4–6)). The least-square estimation of the coefficients of the (order-1) $\lambda(t)$ and (order-2) $\phi(t)$ polynomials of Equation (2) – using the observations of the log-amplitude and phase on the 3 consecutive spectra – turned out to be much less accurate.

5. EXPERIMENTS AND RESULTS

To quantitatively evaluate the precision of the difference method (D) for the estimation of all the model parameters, we ran the same experiments as in [2] and made comparisons to the reassignment method (R) and to the Cramér-Rao bounds (CRB) – again, see [2].

When looking at the results of these experiments (see Figures 1–3), we see that the simple difference method is in fact very good, and compares to the reassignment method.

Regarding the phase, frequency, amplitude, and amplitude modulation, both methods produce very similar and excellent results. In the stationary case (see Figure 2), the simple difference method performs even slightly better. This is also true in case of amplitude modulation only (not shown here to save some space). But when frequency modulation is present (see Figure 3), the performances of both methods degrade in exactly the same way.

The most noticeable differences are for the estimation of the frequency modulation (see Figure 1). The same trend is confirmed: when this frequency modulation is absent, the simple difference method is as good (a) or even better (b). But when the frequency modulation is present, the results are getting worse (c,d) – although acceptable: the difference method exhibits a bias, for very high signal-to-noise ratios (SNRs) though.

6. CONCLUSION AND FUTURE WORK

In this paper, the generalization of the phase vocoder approach to the non-stationary case has been proposed. This difference method gives indeed very good results, comparable to the state of the art. And since it is clearly the simplest method, this makes it a perfect candidate for the implementation in sinusoidal analysis systems. However, there is room for improvement especially regarding the estimation of the frequency modulation when present. Moreover, there is a need for a simple and efficient computation of the Γ_w function for popular analysis windows.

7. ACKNOWLEDGMENTS

The author would like to thank Matthieu Hodgkinson for pointing out [14] and deciding him to finally take some time to investigate the generalization of the difference method to the non-stationary case, with the conclusion that the simplest can also be the best.

This research was partly supported by the French ANR DReaM project (ANR-09-CORD-006).

8. REFERENCES

[1] 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.

[2] Sylvain Marchand and Philippe Depalle, "Generalization of the Derivative Analysis Method to Non-Stationary Sinusoidal Modeling," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Espoo, Finland, September 2008, pp. 281–288.

[3] Brian Hamilton, Philippe Depalle, and Sylvain Marchand, "Theoretical and Practical Comparisons of the Reassignment Method and the Derivative Method for the Estimation of the Frequency Slope," in *Proceedings IEEE WASPAA*, New Paltz, New York, USA, October 2009, pp. 345–348.

[4] Michaël Betser, *Modélisation sinusoïdale et applications à l'indexation audio*, Ph.D. thesis, École Nationale Supérieure des Télécommunications, Paris, France, April 2008.

[5] 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.

[6] Michaël Betser, "Sinusoidal Polynomial Parameter Estimation Using the Distribution Derivative," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4633–4645, December 2009.

[7] Brian Hamilton, "Non-stationary Sinusoidal Parameter Estimation," M.S. thesis, McGill University, Montreal, Quebec, Canada, December 2011.

[8] Brian Hamilton and Philippe Depalle, "A Unified View of Non-Stationary Sinusoidal Parameter Estimation Methods Using Signal Derivatives," in *Proceedings IEEE ICASSP*, Kyoto, Japan, March 2012.

[9] Brian Hamilton and Philippe Depalle, "Comparisons of Parameter Estimation Methods for an Exponential Polynomial Sound Signal Model," in *Proceedings of the AES 45th Conference on Applications of Time-Frequency Processing in Audio*, Helsinki, Finland, March 2012.

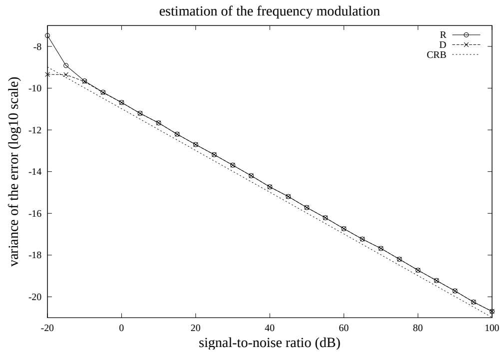
[10] Xue Wen and Mark Sandler, "Notes on Model-Based Non-Stationary Sinusoid Estimation Methods Using Derivatives," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Como, Italy, September 2009.

[11] Sašo Muševič, "Non-Stationary Sinusoidal Analysis," M.S. thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2009.

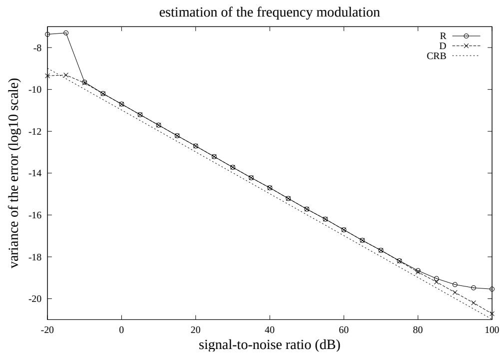
[12] Sašo Muševič and Jordi Bonada, "Comparison of Non-Stationary Sinusoid Estimation Methods Using Reassignment and Derivatives," in *Proceedings of the Sound and Music Computing (SMC) Conference*, Barcelona, Spain, July 2010.

[13] Mototsugu Abe and Julius O. Smith III, "AM/FM Rate Estimation for Time-Varying Sinusoidal Modeling," in *Proceedings IEEE ICASSP*, Philadelphia, New York, USA, March 2005, vol. III, pp. 201–204.

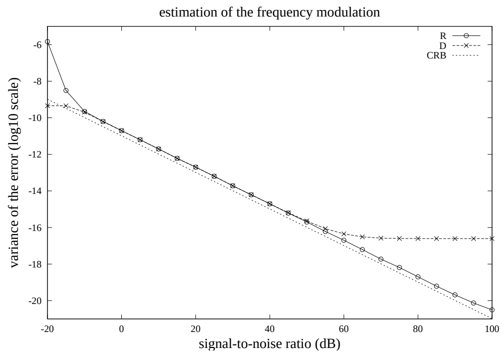
[14] Kevin M. Short and Ricardo A. Garcia, "Signal Analysis Using the Complex Spectral Phase Evolution (CSPE) Method," in *Proceedings of the 120th Convention of the Audio Engineering Society (AES)*, Paris, France, May 2006.



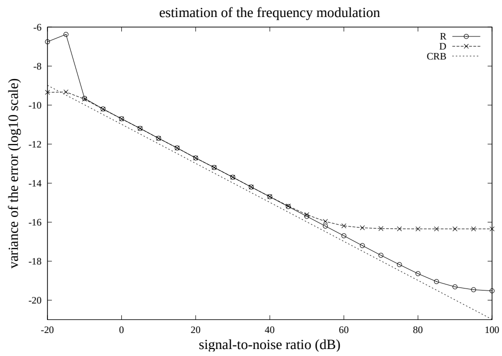
(a) stationary case ($\mu = 0, \psi = 0$)



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

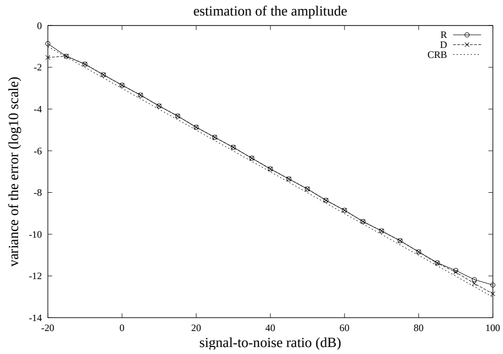


(c) FM-only case ($\mu = 0, |\psi| \leq 10000$)

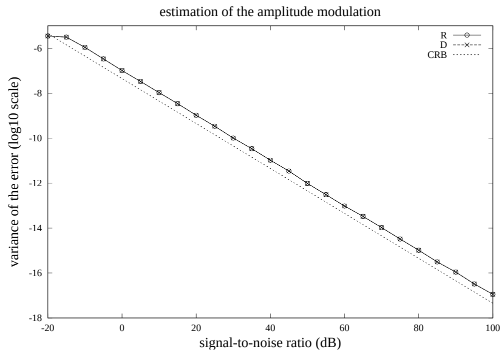


(d) AM/FM case ($|\mu| \leq 100, |\psi| \leq 10000$)

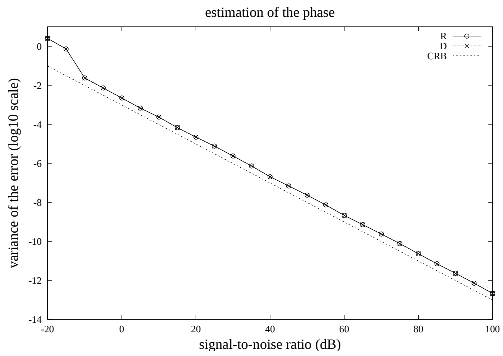
Figure 1: *Frequency modulation estimation error as a function of the SNR (stationary, AM-only, FM-only, and AM/FM cases) for the reassignment (R) and difference (D) methods, and comparison to the Cramér-Rao Bound (CRB).*



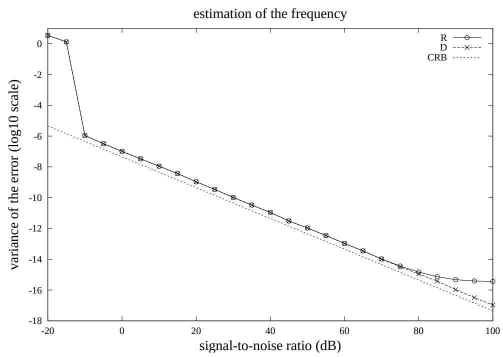
(a)



(b)

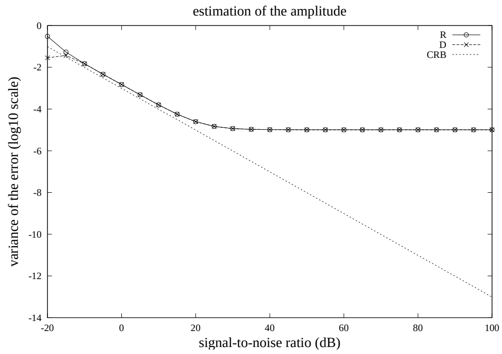


(c)

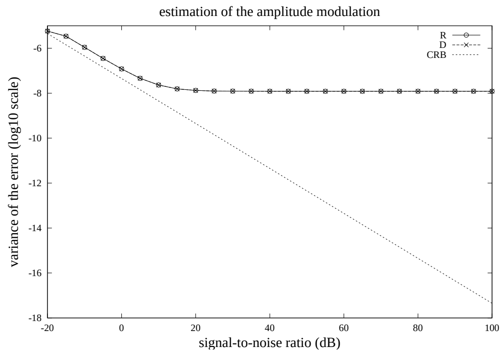


(d)

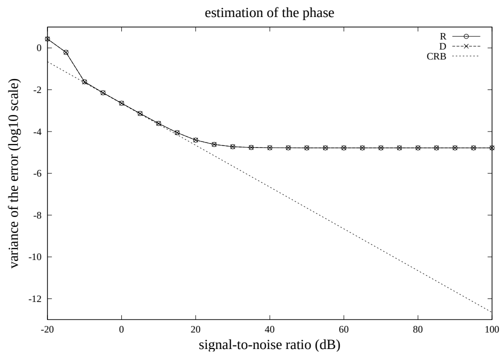
Figure 2: *Estimation errors as functions of the SNR in the stationary case, for the amplitude (a), amplitude modulation (b), phase (c), and frequency (d), with the reassignment (R) and difference (D) methods, and comparison to the Cramér-Rao Bounds (CRB).*



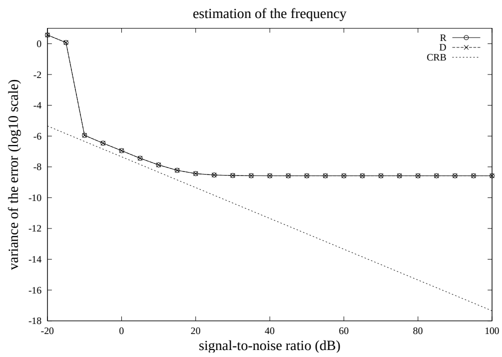
(a)



(b)



(c)



(d)

Figure 3: *Estimation errors as functions of the SNR in the AM/FM case, for the amplitude (a), amplitude modulation (b), phase (c), and frequency (d), with the reassignment (R) and difference (D) methods, and comparison to the Cramér-Rao Bounds (CRB).*