ON THE EQUIVALENCE OF PHASE-BASED METHODS FOR THE ESTIMATION OF INSTANTANEOUS FREQUENCY

Sylvain Marchand and Mathieu Lagrange

SCRIME – LaBRI, University of Bordeaux 1
351 cours de la Libération, F-33405 Talence cedex, France
firstname.name@labri.fr

ABSTRACT

Estimating the frequency of sinusoidal components is a key problem in many applications, such as in sinusoidal sound modeling, where the estimation has to be done with a low complexity, on short-term spectra. Many estimators have therefore been proposed in the literature. Among these, we focus in this paper on a class known as the "phase-based" estimators. Despite their different theoretical backgrounds, we prove that four of these estimators are equivalent, at least in theory. We also demonstrate that these estimators perform roughly similarly in practice, however small differences remain which are mainly due to numerical properties of the mathematical operators used in their implementations.

1. INTRODUCTION

Among numerous other applications, sinusoidal sound modeling requires the estimation of the frequencies of sinusoidal components. The inherent complexity of the analyzed sound signals together with the need for real-time applications often impose the use of methods based on short-term spectra, such as those obtained with the Short-Time Fourier Transform (STFT), often implemented using the Fast Fourier Transform (FFT).

A first class of these estimators considers some values of the magnitude spectrum around a frequency bin to fit a polynomial. The location of the maximum of this polynomial gives the precise frequency of the sinusoidal component [1, 2, 3]. In this paper, we focus on a second class of estimators which explicitly use the phase of the FFT to estimate the frequency: the reassignment method [4], the difference method [5] commonly used in the phase-vocoder approach, and the derivative estimator [6]. This last estimator, originally proposed in [7], has been enhanced in [8] to overcome a loss of precision in the high frequencies underlined in [9]. This enhanced estimator will be named the trigonometric estimator in the remainder of this paper.

After a brief presentation of sinusoidal modeling in Section 2, the four phase-based estimators are reviewed in Section 3 and their theoretical equivalence is demonstrated. First, the equivalence of the reassignment estimator and the derivative one is proven in the case of continuous time. Next, a trigonometric interpretation of the difference estimator and the derivative one demonstrates their equivalence in the discrete case.

The numerical experiments presented in Section 4 globally confirm these results. However, practical implementations of these estimators lead to non negligible differences when compared to the Cramér-Rao Bound (CRB). To evaluate and explain these differences, we first compare these estimators in the case of a complex exponential signal and also in the case of a real sinusoid since real signals are more commonly used in musical applications.

2. SINUSOIDAL MODELING

Additive synthesis is the original spectrum modeling technique. It is rooted in Fourier's theorem, which states that any periodic function can be modeled as a sum of sinusoids at various amplitudes and harmonic frequencies. For stationary pseudo-periodic sounds, these amplitudes and frequencies continuously evolve slowly with time, controlling a set of pseudo-sinusoidal oscillators commonly called partials. This is the well-known McAulay-Quatieri representation [10] for speech signals, also used by Serra [3] in the context of musical signals.

The audio signal s can be calculated from the additive parameters using Equations (1) and (2):

$$s(t) = \sum_{p=1}^{P} A_p \, e^{j\phi_p(t)} \tag{1}$$

$$\phi_p(t) = \phi_p(0) + \int_0^t \omega_p(u) \, du \quad i.e. \quad \omega_p(t) = \frac{d}{dt}\phi_p(t) \tag{2}$$

where P is the number of partials and the parameters of the sinusoidal model are ω_p, A_p, and ϕ_p, which are respectively the instantaneous frequency, amplitude, and phase of the p-th partial.

In the real case, the signal s consists of a sum of (real) sinusoids:

$$s(t) = \sum_{p=1}^{P} A_p \, \cos\left(\phi_p(t)\right). \tag{3}$$

In fact, each sinusoid consists of two complex exponentials, since we have:

$$\cos(x) = \left(e^{+jx} + e^{-jx}\right)/2. \tag{4}$$

The basic method used for estimating the model parameters is the Short-Time Fourier Transform (STFT), where a sliding analysis window w is used to obtain short-term spectra:

$$S_w(\omega,t) = \int_{-\infty}^{+\infty} s(\tau) \, w(\tau - t) \, e^{-j\omega(\tau - t)} \, d\tau. \tag{5}$$

Then the sinusoidal components are searched in each spectrum, using specific estimators.

In this paper, we will focus on the estimation of the frequency, which will be considered as constant during the analysis window w.

3. PHASE-BASED ESTIMATORS

The (short-term) spectra obtained from the STFT with Equation (5) consist of complex values, which in the polar representation are:

$$S(\omega,t) = A(\omega,t)\ e^{j\phi(\omega,t)} \tag{6}$$

and in the present study we will use only the phases:

$$\phi(\omega,t) = \angle S(\omega,t) = \Im(\log(S(\omega,t))) \tag{7}$$

($\angle x$ and $\Im(x)$ denoting respectively the angle and imaginary part of the complex number x).

3.1 Reassignment Method

In usual time-frequency representations, the values obtained when decomposing the signal on the time-frequency atoms are assigned to the geometrical center of the cells (center of the analysis window and bins of the Fourier transform). Auger and Flandrin propose in [4] to assign each value to the center of gravity of the cell's energy. The method uses the knowledge of the analytic first derivative w' of the analysis window w in order to adjust the frequency inside the FFT bin.

More precisely, if we consider Equations (5) and (7), we can compute:

$$\begin{aligned}\hat{\omega} &= \frac{\partial}{\partial t}\phi(\omega,t) = \Im\left(\frac{\partial}{\partial t}\log(S_w(\omega,t))\right)\\ &= \Im\left(\frac{\frac{\partial}{\partial t}\left(\int_{-\infty}^{+\infty} s(\tau)\ w(\tau-t)\ e^{-j\omega(\tau-t)}\ d\tau\right)}{S_w(\omega,t)}\right)\\ &= \Im\left(\frac{j\omega S_w(\omega,t) - S_{w'}(\omega,t)}{S_w(\omega,t)}\right)\end{aligned}$$

that is

$$\hat{\omega} = \omega - \Im\left(\frac{S_{w'}(\omega,t)}{S_w(\omega,t)}\right). \tag{8}$$

In the special case of a single sinusoid ($P = 1$ in Equation (1)), we have $\hat{\omega} = \omega_1$. In the general case of multiple sinusoids, if we consider the spectrum at a frequency ω close to ω_p, we can neglect the influence of the other frequencies, since the analysis window is band limited, and thus we have $\hat{\omega} \approx \omega_p$ for $\omega \approx \omega_p$. The frequency $\hat{\omega}$ gives indeed an excellent estimate of ω_p when estimated at the frequency of the DFT bin nearest to ω_p, where the p-th spectral peak is found in the magnitude spectrum. The reassignment method is the first estimator we consider in our study, or more precisely its discrete version:

$$\hat{\omega} = \frac{k}{N} - \Im\left(\frac{S_{w'}[k,n]}{S_w[k,n]}\right) \tag{9}$$

where k is the DFT bin number of the local maximum in the magnitude spectrum for the sinusoidal component under consideration, and n is the sample index corresponding to the time where this N-point DFT is computed.

3.2 Derivative Method

We have shown in [11, 7] that it is also possible to greatly improve the precision of the classic Fourier analysis by taking advantage of the first d derivatives of the signal itself. For $d = 1$, this method is also known as the derivative algorithm.

3.2.1 Continuous Case

In the case of continuous time, from Equation (1), the analytic derivative of the signal s is given by:

$$\begin{aligned}s'(t) &= \frac{d}{dt}s(t)\\ &= \sum_{p=1}^{P}\frac{d}{dt}A_p\ e^{j\phi_p(t)} + j\sum_{p=1}^{P}A_p\ \frac{d}{dt}\phi_p(t)\ e^{j\phi_p(t)}.\end{aligned}$$

If the amplitudes are constant, their derivatives are zero, and since the derivative of the phases are the frequencies, with Equation (2), we have:

$$s'(t) = j\sum_{p=1}^{P}A_p\ \omega_p\ e^{j\phi_p(t)}. \tag{10}$$

In the special case of a single sinusoid ($P = 1$), we get $s'(t) = j\omega_1 s(t)$, thus $S'(t) = j\omega_1 S(t)$ provided that ω_1 is constant. In the general case of multiple sinusoids, the same considerations as for the reassignment method lead to:

$$S'(\omega,t) = j\omega_p S(\omega,t) \quad \text{for} \quad \omega \approx \omega_p \tag{11}$$

thus we have

$$\hat{\omega} = \Im\left(\frac{S'_w(\omega,t)}{S_w(\omega,t)}\right) \tag{12}$$

which is the continuous version of the derivative method, proposed in [7].

In order to prove the equivalence of this method and the reassignment method, we introduce $\theta = \tau - t$ which gives another expression for the STFT:

$$S_w(\omega,t) = \int_{-\infty}^{+\infty} s(t+\theta)\ w(\theta)\ e^{-j\omega\theta}\ d\theta \tag{13}$$

from which we can derive, as we did for the reassignment to obtain Equation (8):

$$\begin{aligned}\hat{\omega} &= \Im\left(\frac{\partial}{\partial t}\log(S_w(\omega,t))\right)\\ &= \Im\left(\frac{\frac{\partial}{\partial t}\left(\int_{-\infty}^{+\infty} s(t+\theta)\ w(\theta)\ e^{-j\omega\theta}\ d\theta\right)}{S_w(\omega,t)}\right)\\ &= \Im\left(\frac{S'_w(\omega,t)}{S_w(\omega,t)}\right)\end{aligned}$$

that is exactly Equation (12), thus the reassignment and derivative methods are equivalent, at least in theory, since they are two different mathematical formulations of the same physical quantity.

3.2.2 Discrete Case

However, in practice, the signal s is discrete and the derivative of the signal is unknown and must be approximated. In [7], it is proposed to consider the difference s^- as an approximation of the derivative s'. More precisely, we have:

$$s[n] = s(n/F_s) \tag{14}$$

$$s^-[n] = (s[n+1] - s[n])F_s \tag{15}$$

where F_s is the sampling frequency. In fact, s^- defines a low-pass filter of the signal s whose gain is $2F_s \sin(\omega/2)$, and we derived in [6] the derivative estimator in the discrete case:

$$\hat{\omega} = \frac{1}{\pi} \arcsin\left(\frac{S^-[k,n]}{2F_s S[k,n]}\right). \tag{16}$$

However, the following identity:

$$\frac{S[k,n+1]+S[k,n]}{2} - S[k,n] = \frac{S[k,n+1]-S[k,n]}{2}$$

and simple geometric considerations lead us to consider a right-angled triangle, see Figure 1 (lower triangle, delimited by 3 dots). Another of its angles measures $\theta = \Delta_\phi/2$ radians. Simple trigonometric considerations give:

$$\begin{aligned} \cos(\theta) &= \frac{S[k,n+1]+S[k,n]}{2}/S[k,n] \\ \sin(\theta) &= \frac{S[k,n+1]-S[k,n]}{2}/S[k,n] \end{aligned}$$

that is

$$\begin{aligned} \Delta_\phi &= 2\arccos\left(\frac{S[k,n+1]+S[k,n]}{2S[k,n]}\right) \\ \Delta_\phi &= 2\arcsin\left(\frac{S[k,n+1]-S[k,n]}{2S[k,n]}\right). \end{aligned}$$

Considering that the frequency is constant during the time-interval between two successive short-term spectra, with a hop size of H samples, Equation (2) shows that the frequency can be estimated from the phase difference:

$$\hat{\omega} = \frac{1}{2\pi H}\Delta_\phi \tag{17}$$

which gives, for $H = 1$:

$$\hat{\omega}^- = \frac{1}{\pi} \arcsin\left(\frac{S[k,n+1]-S[k,n]}{2S[k,n]}\right) \tag{18}$$

$$\hat{\omega}^+ = \frac{1}{\pi} \arccos\left(\frac{S[k,n+1]+S[k,n]}{2S[k,n]}\right). \tag{19}$$

The $\hat{\omega}^-$ estimator, introduced in [6], will be the second estimator considered in our study. We introduced recently in [8] the $\hat{\omega}^+$ estimator in order to improve the precision of the previous one in the high frequencies. The resulting estimator, called the trigonometric estimator, is the third estimator we consider in this paper:

$$\hat{\omega} = \begin{cases} \hat{\omega}^- & \text{if } k/N < 0.25 \\ \hat{\omega}^+ & \text{otherwise.} \end{cases} \tag{20}$$

3.3 Difference Method

The difference method is used in the classic phase vocoder approach [5]. This is a straightforward application of Equation (17), thus again considering that the frequency is constant during the time-interval between two successive short-term spectra, which is especially the case if they are separated by only 1 sample. The frequency is then estimated directly from the phase difference Δ_ϕ (see Figure 1) by taking care of unwrapping the phase so that this difference is never negative. The resulting estimator, known as the difference estimator, is the fourth and last one we consider in our study:

$$\hat{\omega} = \frac{1}{2\pi} \left(\angle S[k,n+1] - \angle S[k,n]\right)_{\text{unwrap}}. \tag{21}$$

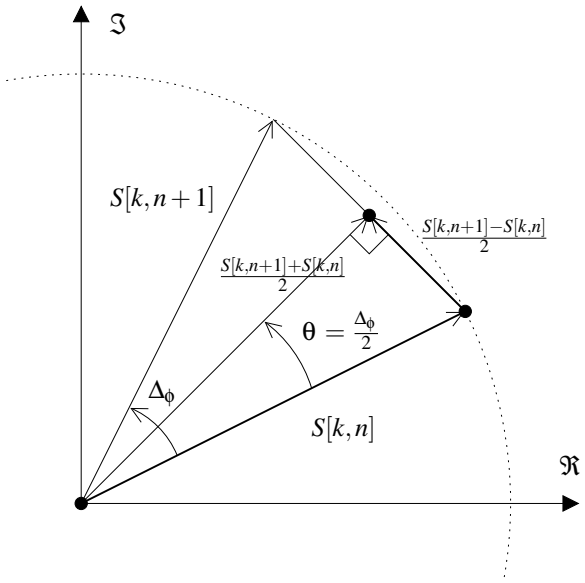


Figure 1: *Vector relationships for phase difference and discrete derivative methods. $S[k,n]$ is the spectrum at DFT bin number k and time index n, and Δ_ϕ is the phase difference between two consecutive (short-term) spectra.*

Figure 1 shows that this estimator is equivalent to the preceding one in the discrete case, at least in theory. Next section will study the relationships among all these estimators on a more practical (implementation) point of view.

4. PRACTICAL EXPERIMENTS

To compare the estimators reviewed in the previous section, we consider in turn a complex and a real signal composed of a periodic part x with amplitude unity and constant frequency embedded in noise y. The power of the noise is chosen to achieve a desired Signal-to-Noise Ratio (SNR) expressed in decibels:

$$\text{SNR} = 10\log_{10}\left(\frac{\text{var}(x)}{\text{var}(y)}\right). \tag{22}$$

In the experiments, the SNR ranges from -20 dB to 100 dB.

We use frames of $N = 128$ samples ($F_s = 4$ kHz) and consider 400 different (normalized) frequencies ranging from 0 to 0.5. These bounds are exclusive, so that the first evaluated frequency is 0.0025. For each frequency, 30 different phases are evaluated from 0 to 2π. At each evaluation, the noise is randomized. The (periodic) Hann window is used implicitly prior to any DFT. For all the tested methods, the detection is operated by picking the greatest local maximum in the power spectrum.

4.1 Complex Case

When evaluating the performance of an estimator in terms of variance of the estimation error, an interesting element

to compare with is the Cramér-Rao Bound (CRB). This bound is defined as the limit to the best possible performance achievable by an estimator given a dataset.

Let us consider a complex exponential x (of amplitude 1) in a Gaussian complex noise y:

$$x[n] = e^{j2\pi\omega n+\Phi} \quad \text{and} \quad y[n] = 10^{-\mathrm{SNR}/20} z[n] \tag{23}$$

where ω is the normalized frequency and z is a Gaussian noise of variance 1. The variance of the signal part x is 1, and the variance of the noise part y is $\mathrm{var}(y) = \sigma^2 = 10^{-\mathrm{SNR}/10}$. The analyzed signal is $s = x + y$. For the case of the estimation of the frequency ω of a complex exponential in noise, the lower Cramér-Rao bound is [12]:

$$\mathrm{CRB}_c = \frac{6\sigma^2}{a^2 N(N^2-1)} = \frac{6}{N(N^2-1)}\, 10^{-\mathrm{SNR}/10} \tag{24}$$

where a is the amplitude of the exponential (here $a = 1$), and the SNR is given by Equation (22). We can easily show that, in the log scales, the CRB in function of the SNR is a line of slope -1.

As asserted by the theoretical derivations of Section 3, most of the methods perform similarly, see Figure 2(a). More precisely, the difference estimator performs slightly better than the trigonometric estimator. A loss of performance at high SNR can be observed for the reassignment due to a bias of the frequency estimate [9]. As as far as the entire frequency range is concerned, the derivative method performs badly due to a loss of precision in the high frequencies, see Figure 2(b). This loss of precision can be explained by the numerical imprecision of the arcsin function of Equation (16) when the argument is close to zero, as shown in [8].

4.2 Real Case

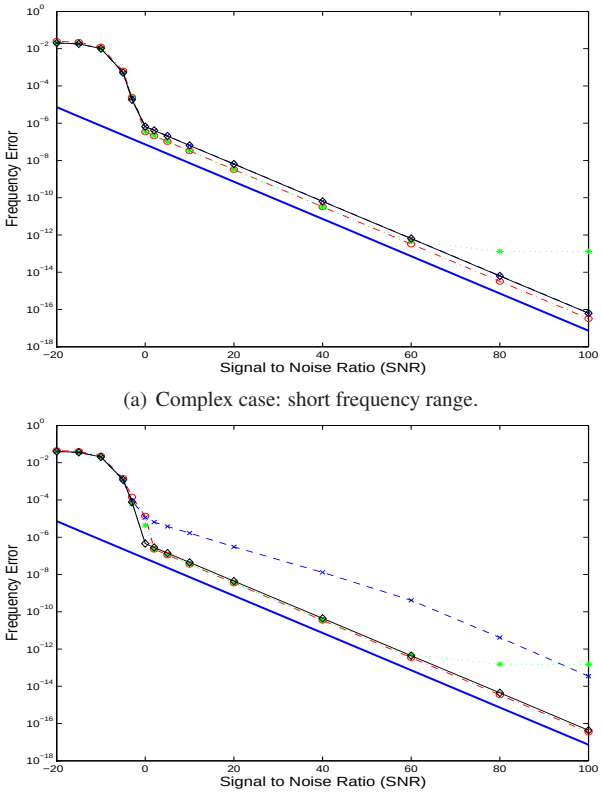
Musical applications consider real sinusoids rather than complex exponentials. We then consider in this section the signal $s = x + y$ made of a sinusoid x (of amplitude 1) in a Gaussian noise y:

$$x[n] = \sin(2\pi\omega n + \Phi) \text{ and } y[n] = \frac{1}{\sqrt{2}} 10^{-\mathrm{SNR}/20} z[n]. \tag{25}$$

We use the $1/\sqrt{2}$ normalizing factor to ensure the validity of Equation (22), because in the real case the variance of the sinusoid is $1/2$, and we still consider, by definition, $\mathrm{var}(y) = \sigma^2$. Therefore, for the case of the estimation of the frequency ω of a real sinusoid in noise, the lower Cramér-Rao bound is shown to be twice the CRB in the complex case ($\mathrm{CRB}_r = 2\,\mathrm{CRB}_c$), see [13].

The spectrum of a real sinusoid is made of two Dirac's impulses, one located at frequency ω and the other at $-\omega$, see Equation (4), and the spectrum of the sampled signal is F_s-periodic. As a consequence, the more the frequency of the analyzed sinusoid is close to 0 or $F_s/2$, the more the interference between the two complex exponentials is pronounced. This can greatly disturb the estimators, thus changing their relative performances in the real case.

Therefore, when the limited frequency range is considered (see above), the results are equivalent to the complex case, see Figure 3(a). If the whole frequency range is considered, the performances are limited by the interference phenomenon, so that the squared error is held asymptotically constant at SNR higher than 10 dB, see Figure 3(b).



(a) Complex case: short frequency range.

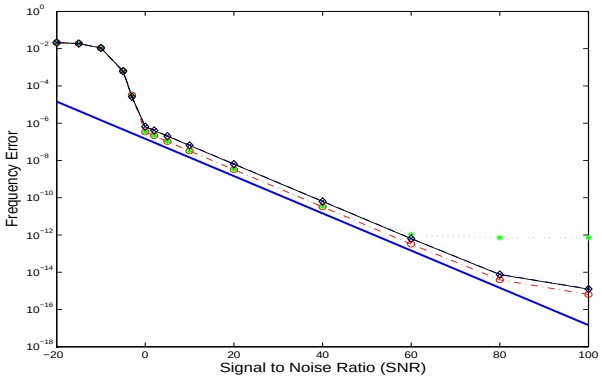
(b) Complex case: entire frequency range.

Figure 2: *Performance of the tested estimators for the analysis of a complex exponential with frequency lying in the* $]0.24, 0.25[$ *normalized frequency range (a), and in the* $]0, 0.5[$ *range (b): the reassignment method (dotted line with* $*$*), the difference estimator (dash-dotted line with* $\circ$*), the derivative estimator (dashed line with* $\times$*), and the trigonometric estimator (solid line with* $\diamond$*). The CRB is plotted with a double solid line.*

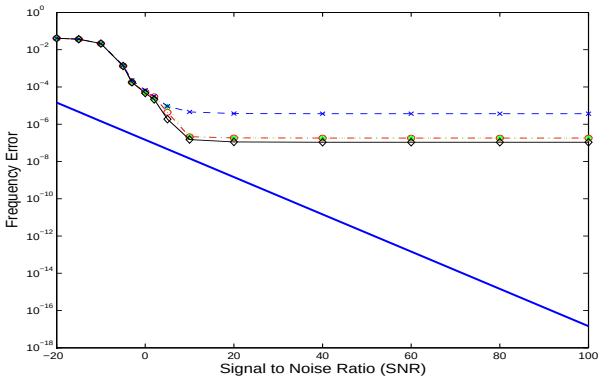
We see that the difference and reassignment methods perform roughly equally, slightly worse than the trigonometric estimator. This is due to the fact that the first DFT bin $S[0]$ is by definition purely real when the signal is real. Consequently, the difference and reassignment estimators will always estimate a frequency zero if the frequency of the analyzed sinusoid falls into this bin, *i.e.* $\omega < 1/N$, see Figure 4.

5. CONCLUSION

In this article, we have demonstrated the theoretical equivalence of phase-based estimators. This equivalence was guessed in practice by Keiler and Marchand in [14], and partly demonstrated in theory by Hainsworth [15] during his Ph.D. Numerical experiments confirmed this result although slight differences were observed, which were mainly due to numerical imprecision. The reassignment estimator should therefore be avoided due to a loss of precision at high SNR and the trigonometric method should be preferred when considering real signals with low-frequency components.



(a) Real case: short frequency range.



(b) Real case: entire frequency range.

Figure 3: *Performance of the tested estimators for the analysis of a real sinusoidal signal with frequency lying in the* $]0.24, 0.25[$ *normalized frequency range (a), and in the* $]0, 0.5[$ *range (b): the reassignment method (dotted line with* $*$*), the difference estimator (dash-dotted line with* $\circ$*), the derivative estimator (dashed line with* $\times$*), and the trigonometric estimator (solid line with* $\diamond$*). The CRB is plotted with a double solid line.*

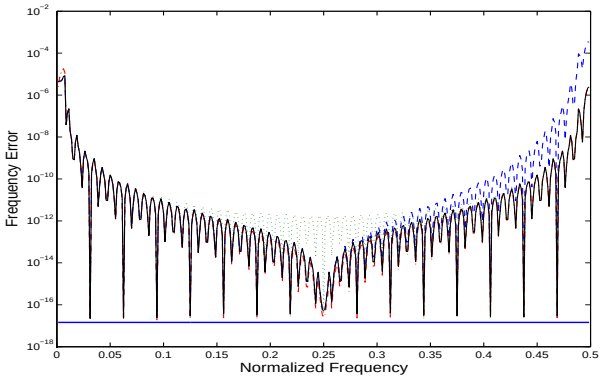


Figure 4: *Performance of the tested estimators at SNR*=100 *dB versus the frequency of the analyzed sinusoid: the reassignment method (dotted line), the difference estimator (dash-dotted line), the derivative estimator (dashed line), and the trigonometric estimator (solid line). The CRB is plotted with a double solid line.*

REFERENCES

[1] Thomas Grandke, "Interpolation Algorithms for Discrete Fourier Transforms of Weighted Signals," *IEEE Trans. on Instruments and Measurements*, vol. 32, no. 2, pp. 350–355, 1983.

[2] Malcolm D. Macleod, "Fast Nearly ML Estimation of the Parameters of Real or Complex Single Tones or Resolved Multiple Tones," *IEEE Trans. on Signal Processing*, vol. 46, no. 1, pp. 141–148, 1998.

[3] Xavier Serra, *Musical Signal Processing*, chapter Musical Sound Modeling with Sinusoids plus Noise, pp. 91–122, Studies on New Music Research. Swets & Zeitlinger, Lisse, the Netherlands, 1997.

[4] François Auger and Patrick Flandrin, "Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method," *IEEE Trans. on Signal Processing*, vol. 43, pp. 1068–1089, 1995.

[5] Daniel Arfib, Florian Keiler, and Udo Zölzer, *DAFx – Digital Audio Effects*, J. Wiley and Sons, 2002, chapter 9, pp. 299–372.

[6] Sylvain Marchand, *Sound Models for Computer Music (Analysis, Transformation, and Synthesis of Musical Sound)*, Ph.D. thesis, University of Bordeaux 1, France, LaBRI, 2000.

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

[8] Mathieu Lagrange, Sylvain Marchand, and Jean-Bernard Rault, "Improving Sinusoidal Frequency Estimation Using a Trigonometric Approach," in *Proc. DAFx Conference*, Madrid, 2005, pp. 141–146.

[9] Stephen W. Hainsworth and Malcolm D. Macleod, "On Sinusoidal Parameter Estimation," in *Proc. DAFx Conference*, London, 2003, pp. 151–156.

[10] Robert J. McAulay and Thomas F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[11] Sylvain Marchand, "Improving Spectral Analysis Precision with an Enhanced Phase Vocoder Using Signal Derivatives," in *Proc. DAFx Conference*, Barcelona, 1998, pp. 114–118.

[12] David C. Rife and Robert R. Boorstyn, "Single-Tone Parameter Estimation from Discrete-Time Observations," *IEEE Trans. on Information Theory*, vol. IT-20, pp. 591–598, 1974.

[13] Steven M. Kay, *Fundamentals of Statistical Signal Processing – Estimation Theory*, Signal Processing Series. Prentice Hall, 1993.

[14] Florian Keiler and Sylvain Marchand, "Survey on Extraction of Sinusoids in Stationary Sounds," in *Proc. DAFx Conference*, Hamburg, 2002, pp. 51–58.

[15] Stephen W. Hainsworth, *Techniques for the Automated Analysis of Musical Audio*, Ph.D. thesis, University of Cambridge, United Kingdom, Signal Processing Group, Department of Engineering, 2003.