Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

Review and discussion on classical STFT-based frequency estimators

Michaël Betser[1], Patrice Collen[1], Gaël Richard[2], Bertrand David[2]

[1] *France Telecom R&D, Cesson-Sévigné, 35512, France*

[2] *Telecom Paris, Paris, 75634, France*

Correspondence should be addressed to Michaël Betser (michael.betser@francetelecom.com)

ABSTRACT

Sinusoidal modeling is based on the decomposition of audio signals into a sum of sinusoidal components plus a noise residual part. It involves accurate sinusoid parameters estimation and, in particular, accurate frequency estimation. A broad category of methods uses the Fast Fourier Transform (FFT) as a starting point to compute frequency. All these methods present very similar forms of estimators, but the relations between them are not yet fully understood. This work proposes to take a deeper look into these relations. The first goal of this work is to present a clear review and description of the classical FFT-based frequency estimators. A new estimator similar to the phase vocoder is presented. The second goal of this work is to identify the common hypotheses and the common steps of the processes for this category of estimators. Lastly, experimental comparisons are given.

1. INTRODUCTION

Sinusoidal modeling [1] is a very popular and efficient representation for speech and music signals. A number of related applications in coding [2], analysis/synthesis [3, 4] or sound effect processing bear witness to its popularity. This work adresses the estimation issues that all the associated methods have to cope with.

Many frequency estimators use the Short-Time Fourier Transform (STFT) as a starting point. Since the STFT is parametrized in terms of analysis time index t_m, frequency bin ω_k and analysis window h, a typology can be derived according to whether the derivation of the estimator uses the transform evaluated at different instants t_m, different frequencies ω_k or different windows. To the first category belongs the phase-vocoder [5] and, more generally, the phase-derivative-based algorithms [6]. The

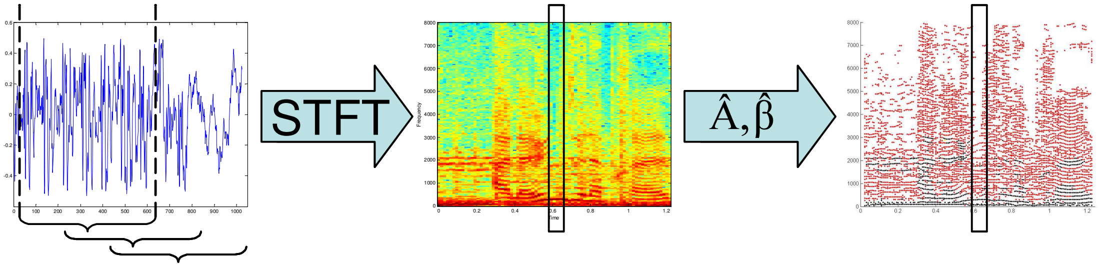


Fig. 1: FFT-based analysis method

amplitude or phase spectrum interpolation methods [7, 8, 9] can be attached to the second one while, for some specific implementation, the reassigned spectrogram falls into the last category [10]. A comparison of these state-of-the-art frequency methods can be found in [11, 12].

The first purpose of this article is to present a review of these methods and, during the review, two new estimators will be described, one similar to the derivative method, and one similar to the DFS interpolators using phase. For each estimator, the conditions of application, the formula and the precise algorithm used in practice will be given. The second purpose is to highlight the similarities between the STFT-based frequency estimators. It will be shown that all the methods presented in this work are actually based on a DFT ratio. In order to emphasize this remark, the ratio used by each method will be clearly defined.

The outline of the paper is the following: a short description of the general structure of the methods is given in section 2, before a presentation in section 3 of the two sinusoidal models on which these estimators are based, namely the linear phase model and the quadratic phase model. Then, a wide variety of estimators is described in detail, section 4. In section 5, the common principles of these estimators are discussed. Experimental comparisons end the main part of the document, section 6, preceding a conclusion note.

2. SYSTEM OVERVIEW

All the estimators presented and compared here use the scheme illustrated on figure 1. The first step is the STFT: the signal is analyzed using a sliding window and a Discrete Fourier Transform (DFT) is performed on the windowed signal. In the second step, a combination of the STFT time-frequency points leads to the estimation of the sinus parameters. Based on this scheme, all the methods presented in this work use a unique STFT, which means that all the methods compared have the same order of complexity.

In this work, the STFT is defined as:

$$X(t_m, \omega_k; h) \triangleq \sum_{n=0}^{N-1} x(\tau_n + t_m)\, h(\tau_n)\, \mathrm{e}^{-j\,\tau_n \omega_k} \quad (2.1)$$

where N is the size in samples of the window support h, F is the sampling frequency, k is the frequency bin, $\tau_n \triangleq n/F$ is the time in seconds of the corresponding sample number n. $t_m \triangleq m/F$ is the beginning time of the STFT window. Finally, $\omega_k \triangleq \frac{2\pi k F}{P}$ is the pulsation of the bin k. The length of the STFT P and the length of the window N are supposed to be equal: $P = N$.

During the second step, the estimation is usually performed using maximum bins. A maximum bin k at the instant t_m is defined as:

$$|X(t_m, \omega_{k-1}; h)| < |X(t_m, \omega_k; h)|$$
$$|X(t_m, \omega_{k+1}; h)| > |X(t_m, \omega_k; h)|$$

Other definitions of maximum bins are possible, but are beyond the scope of this article.

3. SIGNAL MODEL

Our purpose is to study the evolution of partial parameters. A partial is an oscillator whose amplitude and frequency may vary upon time:

$$x(t) \triangleq A(t)\, \mathrm{e}^{j\phi(t)} \quad (3.1)$$

where A denotes the amplitude and ϕ the phase. In this article, only the case of a complex sinusoid (cisoid) is considered, but with no loss of generality because the problem of the real sinusoid is usually reduced to the cisoid problem.

The model defined in equation (3.1) is very general so hypotheses are usually added to the model, on amplitude and on frequency. From these hypotheses a local model, valid in the neighborhood of a time t, will be defined. In literature, two models can be found.

The first one, the linear phase model, supposes a quasi-constant frequency and quasi-constant amplitude, which corresponds to the local model:

$$x_1(\tau) \triangleq A\,\mathrm{e}^{j(\alpha+\beta\tau)} \tag{3.2}$$

where α is the initial phase and β the frequency[1]. This model is the most widely used. It is the base for the phase vocoder estimation [5], for the derivative algorithm [13], for amplitude spectrum interpolation [7] and for amplitude spectrum interpolation using phase [14, 8].

The second model is the quadratic phase model. It supposes a quasi-linear frequency (quadratic phase) and a quasi-constant amplitude:

$$x_2(\tau) \triangleq A\,\mathrm{e}^{j(\alpha+\beta\tau+\gamma\tau^2/2)} \tag{3.3}$$

where γ is the frequency slope. This model is much less used. Frequency reassignment [10] is known for localizing perfectly chirp signals. So one can say that it is the model supposed in the frequency reassignment method.

When considering the quadratic phase model, a frequency estimation $\hat{\beta}$ is made for a particular time $\hat{t}$ which must be specified. In a constant frequency model, there is no particular time for the estimation as the frequency is supposed to be constant.

4. DESCRIPTION OF THE ESTIMATORS

4.1. Phase-vocoder-based methods

The term "vocoder", derived from "voice coder", originally refers to a speech analyzer and synthesizer. This method of analysis/synthesis has become famous for its ability to modify sound and its use as an electronic musical instrument.

The phase vocoder is a particular vocoder which uses a representation based on short-time amplitude and phase spectra [15]. It was the very first method to be used in trying to extract the sinusoidal parameters in an analysis step. Portnoff [16] demonstrated an efficient method of building the required filter banks digitally, using the FFT algorithm, which allowed a broad number of real-time applications. The heart of the method is the frequency computation, which is done by using a discrete derivation of the phase, as an approximation of the continuous definition of frequency[2].

All the phase vocoder methods suppose a linear phase model.

Short-term phase vocoder

The frequency is, by definition, the derivative of the phase function. For a single cisoid with linear or constant frequency, the continuous derivative can be replaced by the discrete derivative without any approximation:

$$\beta = \frac{\phi(t_2) - \phi(t_1)}{T} \tag{4.1}$$

where T is the time length between the two measured phases used to compute the frequency. When measured in samples, this interval is called hop size. The problem is reduced to a phase estimation problem: if the phases in t_1 and in t_2 are perfectly known, the frequency will be perfectly computed. The usual phase vocoder estimation of frequency makes use of the identity between the Fourier phase and the cisoid phase, when the frequency is constant:

$$\begin{aligned} \mathcal{H}_v &\triangleq \frac{X(t_2, \omega_k; h)}{X(t_1, \omega_k; h)} && (4.2) \\ \hat{\beta} &= \frac{\arg(\mathcal{H}_v)}{T} && (4.3) \end{aligned}$$

The first phase vocoders, as the one described in [15], used two phases computed at one sample interval ($T = 1/F$). One frequency estimation required two

[1] The terms 'frequency' and 'pulsation' are used equally in this article. It is a misuse of language but the term 'frequency' is more suggestive than 'pulsation'.

[2] Another more restricted meaning for the term phase vocoder refers only to this particular frequency estimator.

Fourier transforms or FFT. But further developments showed that for one sample interval, the second FFT, could be avoided for some windows, such as the rectangular and the Hann windows [17].

Algorithm: One sample vocoder (1SV)

1. Compute the FFTs for a time t_1 and $t_2 = t_1 + 1/F$
2. For each maximum bin k at the time t_1
3. Compute $\mathcal{H}_v$
4. Compute $\hat{\beta} = \frac{\arg(\mathcal{H}_v)}{T}$

Long-term phase vocoder

A larger time interval T has also been used, in order to improve the frequency estimation for constant frequency signals [18, 5]. However, this time could not be increased indefinitely as the constant frequency model had to be a local model in order to analyze real signals. Another problem came from the phase indetermination for larger time intervals. Indeed, for a given pulsation ω, the phase increment between the two measures is ωT. If T is large enough, the increment can be greater than 2π. But all the phases computed via the DFTs are inside the interval $[0, 2\pi[$ which causes phase indetermination.

$$\hat{\beta} = \frac{\arg(\mathcal{H}_v) + 2\pi n}{T} \tag{4.4}$$

where n is an integer, which has to be determined, corresponding to the number of periods the phase increased. Estimation of n can be done by unwrapping the phase. It consists in choosing a frequency of reference Ω such that the difference between the phase increase for the frequency Ω and the phase increase for the true frequency β is not more than one period:

$$|\Omega - \beta| < \frac{\pi}{T} \tag{4.5}$$

A practical estimator for n is given by McAulay in [1]:

$$\hat{n} = \text{round}((-\arg(\mathcal{H}_v) + .5(\omega_1 + \omega_2)T)/(2\pi)) \tag{4.6}$$

Here, the frequency reference is $\Omega = (\omega_1 + \omega_2)/2$. For a sinusoid with constant frequency, and from the definition of $\mathcal{H}_v$, the frequency ω_k can be used as a reference: $\omega_1 = \omega_2 = \omega_k$.

Algorithm: Long term vocoder (LV)

1. Compute the FFTs for a time t_1 and $t_2 = t_1 + T$
2. For each maximum bin k at the time t_1
3. Compute $\mathcal{H}_v$ and $\hat{n}$.
4. Compute $\hat{\beta} = \frac{\arg(\mathcal{H}_v + 2\pi\hat{n})}{T}$

4.2. Derivative algorithm

The derivative method has been introduced by Marchand [13]. This method presents a strong similarity with the phase vocoder method: the model used is the linear phase model and they are both computed in practice for the same frequency bin with a hop size T. The ratio $\mathcal{H}$ for the derivative method is the following:

$$\mathcal{H}_d \triangleq \frac{X(t_2, \omega; h) - X(t_1, \omega; h)}{X(t_1, \omega; h)} \tag{4.7}$$

Here, we suppose that the frequency is constant, which is the usual hypothesis of the derivative method. Remark that $\mathcal{H}_d = \mathcal{H}_v - 1$ and that $|\mathcal{H}_v| = 1$ for a constant frequency:

$$\begin{aligned} \mathcal{H}_v &= \cos(\arg(\mathcal{H}_v)) + j\sin(\arg(\mathcal{H}_v)) \\ |\mathcal{H}_d|^2 &= 2(1 - \cos(\arg(\mathcal{H}_v))) \\ &= 4\sin^2(\frac{\arg(\mathcal{H}_v)}{2}) \\ \arg(\mathcal{H}_v) &= 2\arcsin(\frac{|\mathcal{H}_d|}{2}) \end{aligned}$$

This last relation combined with equation (4.3) leads to the derivative method:

$$\hat{\beta} = \frac{2}{T}\arcsin(\frac{|\mathcal{H}_d|}{2}) \tag{4.8}$$

This estimator is known to become unstable for high frequencies, as the argument of the arcsin tends to 1 [19]. A small error on the argument of the arcsin for high frequencies will cause a large error of estimation.

Arccos estimator

The arccos estimator is presented in Lagrange's thesis [19]. Consider the following ratio:

$$\mathcal{H}_c \triangleq \frac{X(t_2, \omega_k; h) + X(t_1, \omega_k; h)}{X(t_1, \omega_k; h)} \tag{4.9}$$

In a similar way to the previous demonstration, one can show that:

$$\hat{\beta} = \frac{2}{T} \arccos\left(\frac{|\mathcal{H}_c|}{2}\right) \tag{4.10}$$

This estimator presents a symmetrical problem compared to the arcsin estimator: as the frequency tends to zero, the argument of the arccos will tend to 1, leading to an instability of the estimator [19].

Arctan estimator

We now introduce a new estimator similar to the previous methods but avoiding their instability as this estimator is based on an arctan function. This estimator is derived from the ratio:

$$\mathcal{H}_t \triangleq \frac{X(t_2, \omega_k; h) - X(t_1, \omega_k; h)}{X(t_1, \omega_k; h) + X(t_2, \omega_k; h)} \tag{4.11}$$

Supposing the linear phase model, we know that:

$$|\mathcal{H}_d|^2 = 4\sin^2\left(\frac{\arg(\mathcal{H}_v)}{2}\right)$$
$$|\mathcal{H}_c|^2 = 4\cos^2\left(\frac{\arg(\mathcal{H}_v)}{2}\right)$$

Then,

$$\left|\frac{\mathcal{H}_d}{\mathcal{H}_c}\right|^2 = \tan^2\left(\frac{\arg(\mathcal{H}_v)}{2}\right)$$

Using equations (4.7) and (4.9), $|\frac{\mathcal{H}_d}{\mathcal{H}_c}| = |\mathcal{H}_t|$, and a tangent estimator can be derived:

$$\hat{\beta} = \frac{2}{T} \arctan\big(|\mathcal{H}_t|\big) \tag{4.12}$$

To these three estimators correspond the three possibilities for computing the same angle. The arctan estimator seems very similar to the phase vocoder, and experimentally they give very close results (cf. section 6). The algorithm for these three methods is identical to the phase vocoder algorithm and is given only for the derivative method:

Algorithm: Derivative method (D)

1. Compute the FFTs for a time t_1 and $t_2 = t_1 + 1/F$
2. For each maximum bin k at the time t_1
3. Compute $\mathcal{H}_d$
4. Compute $\hat{\beta} = \frac{2}{T} \arcsin\big(\frac{|\mathcal{H}_d|}{2}\big)$

4.3. Spectrum interpolators

A DFS interpolator, is a frequency estimator using discrete spectrum points. All the methods presented here use a linear phase model.

Without phase

The estimator presented in this section proposes to interpolate the magnitude spectrum (or the log magnitude spectrum) using a curve model. The frequency response of the window used must be coherent with the function chosen. Numerous DFS interpolators exist. We have chosen to retain only the parabolic interpolation, which is used most frequently.

Parabolic interpolation [11, 7] tries to interpolate the magnitude spectrum using a parabola. Several windows can be used with this method, like the Hann window or the truncated Gaussian window, which are approximately parabolic near the peak maximum, in db scale. Zero-padding is often used with this method. Noting $R_1(m) = |X(t, \omega_{k+m}; h)|$, this estimator is:

$$\mathcal{H}_{psi} \triangleq \frac{R_1(-1) - R_1(1)}{R_1(-1) - 2R_1(0) + R_1(1)} \tag{4.13}$$
$$\hat{\beta} = \omega_k + \pi \mathcal{H}_{psi} F/P \tag{4.14}$$

Here, the size P of the Fourier transform should be advantageously superior to the size N of the window (zero-padding). The performance of the method strongly depends on the window used and other more adapted windows can be designed. A complete study of the influence of the parameters (i.e. window type, window length, zero-padding factor) on the performance can be found in [7].

Algorithm: Parabolic spectrum interpolation (PSI)

1. Compute the FFT for a time t
2. For each maximum bin k at the time t
3. Compute $\mathcal{H}_{psi}$
4. Compute $\hat{\beta} = \omega_k + \pi \mathcal{H}_{psi} F/P$

With phase

Recent works have proposed to use the complex spectrum instead of modulus spectrum as in usual interpolation methods. The idea is to introduce the phase information in the interpolation. But this method differs notably from the other interpolation methods, as it does not try to estimate the parameters of an interpolating function. The resulting estimators have more in common with the preceding methods, like the phase vocoder. The DFS interpolators using the complex spectrum have been introduced by Barry Quinn in [14], and a generalization of this method can be found in [8]. A comparison with other classical estimators can be found in [12].

The estimators presented in [14, 8] make the hypothesis of a rectangular window $h_{rec}(t) = 1$, because the discrete Fourier transform of a cisoid $x_1(t) = A\,\mathrm{e}^{j(\alpha+\beta t)}$ with a rectangular window has a simple analytical expression. Using the framework proposed by Quinn [14] and Macleod [8], many estimators can be proposed. Macleod's 3-sample estimator has been retained, because of the interesting performances given in [8]. Noting [3] $R_2(m) = \Re(X(t,\omega_{k+m};h_{rec})X(t,\omega_k;h_{rec})^*)$

$$\mathcal{H}_{m3} \triangleq \frac{R_2(-1) - R_2(1)}{2R_2(0) + R_2(-1) + R_2(1)} \tag{4.15}$$

$$\hat{\beta} = \omega_k + 2\pi F/P \frac{(\sqrt{1 + 8\mathcal{H}_{m3}^2} - 1)}{4\mathcal{H}_{m3}} \tag{4.16}$$

Algorithm: Macleod 3-sample estimator (M3)

1. Compute the FFT for a time t
2. For each maximum bin k at the time t
3. Compute $\mathcal{H}_{m3}$
4. Compute $\hat{\beta} = \omega_k + 2\pi F/P \frac{(\sqrt{1+8\mathcal{H}_{m3}^2}-1)}{4\mathcal{H}_{m3}}$

In a previous work [20], a similar estimator has been presented, which allows the use of non rectangular windows. The formulation of the estimator is slightly different from [20] because the definition of the STFT in [20] is the zero-phased STFT.

$$\mathcal{H}_F \triangleq \frac{\lambda X(t,\omega_k;h) - \lambda^* X(t,\omega_{k'};h)}{\lambda X(t,\omega_k;h) + \lambda^* X(t,\omega_{k'};h)} \tag{4.17}$$

$$\hat{\beta} = \omega_b + \mathrm{sign}(k - k^{'})\Re(\mathcal{H}_F)\frac{\Gamma(h_c)}{\Gamma(\tau.h_s)} \tag{4.18}$$

where $\lambda \triangleq \mathrm{e}^{-j\pi(N-1)/(2P)}$ is a parameter used to zero-phase the Fourier transforms, $\tau_M \triangleq (N-1)/(2F)$ is the time corresponding to the middle of the window, $\omega_b \triangleq (\omega_k + \omega_{k'})/2$ is the mean frequency of the bins used, and h_s and h_c are new analysis windows derived from h:

$$h_s(\tau) \triangleq \sin(\pi F(\tau - \tau_M)/P).h(\tau),$$
$$h_c(\tau) \triangleq \cos(\pi F(\tau - \tau_M)/P).h(\tau)$$

At last, $\Gamma(h)$ is the sum function of the elements of the window:

$$\Gamma(h) \triangleq \sum_{n=0}^{N-1} h(\tau_n) \tag{4.19}$$

$\Gamma(h_c)$ and $\Gamma(\tau.h_s)$ are two parameters which can be computed in advance.

The demonstration of the method is in two steps. First, the ratio $\mathcal{H}_F$ is shown to be equal to:

$$\mathcal{H}_F = j\,\frac{X(t,\omega_b;h_s)}{X(t,\omega_b;h_c)} \tag{4.20}$$

Then using the definition of x_1 (3.2), a first order Taylor expansion around the frequency β of $X(t,\omega_b;h_s)$ $X(t,\omega_b;h_c)$ is done, leading to the formula (4.18).

Algorithm: Two-sample estimator with window (F)

1. Initialization: compute $\Gamma(h_c)$ and $\Gamma(\tau.h_s)$.
2. Compute the FFT for a time t
3. For each maximum bin k at the time t
4. Select the second maximum $k' = \arg\max_{i\in\{k+1,k-1\}} |X(t,\omega_i;h)|$
5. Compute $\mathcal{H}_F$
6. Compute $\hat{\beta} = \omega_b + \mathrm{sign}(k - k^{'})\Re(\mathcal{H}_F)\frac{\Gamma(h_c)}{\Gamma(\tau.h_s)}$

[3] In this article, $\Re()$ denotes the real part and $\Im()$ the imaginary part of a complex number.

4.4. Frequency reassignment

With a frequency formulation, the justification of the reassignment method for the quadratic phase is very short. Let f be $f(\tau) \triangleq x_2(\tau).h(\tau)$. Using (3.3), we have:

$$\begin{aligned}\frac{df}{d\tau}(\tau) =& \frac{dx_2}{d\tau}(\tau)h(\tau) + \frac{dh}{d\tau}(\tau)x_2(\tau)\\ =& j(\gamma\tau + \beta)x_2(\tau)h(\tau) + \frac{dh}{d\tau}(\tau)x_2(\tau) \quad (4.21)\end{aligned}$$

The continuous Fourier transform (FT) is defined by:

$$FT(x;\omega) \triangleq \int_{-\infty}^{+\infty} x(\tau)\,\mathrm{e}^{-j\omega\tau}\,d\tau \qquad (4.22)$$

Applying the FT on the relation (4.21) leads to:

$$\begin{aligned} j\omega FT(f;\omega) &= j\gamma FT(\tau f;\omega) + j\beta FT(f;\omega)\\ &\quad + FT(\frac{\partial h}{\partial \tau}x_2;\omega)\\ \Leftrightarrow \beta + \gamma\frac{FT(\tau f;\omega)}{FT(f;\omega)} &= \omega + j\frac{FT(\frac{\partial h}{\partial \tau}x_2;\omega)}{FT(f;\omega)}\\ \Rightarrow \beta + \gamma\Re\Big(\frac{FT(\tau f;\omega)}{FT(f;\omega)}\Big) &= \omega - \Im\Big(\frac{FT(\frac{\partial h}{\partial \tau}x_2;\omega)}{FT(f;\omega)}\Big) \end{aligned} \qquad (4.23)$$

The usual formulation of the reassignment uses $\mathcal{D}h(\tau) \triangleq \frac{\partial h}{\partial \tau}(\tau)$ and $\mathcal{T}h(\tau) \triangleq \tau.h(\tau)$. The first member of equation (4.23) is the frequency of the partial for the time $\hat{t} = t + \Re(\frac{FT(\mathcal{T}h.x_2;\omega)}{FT(h.x_2;\omega)})$, which is the time reassignment operator. The second part of the equation corresponds to the frequency reassignment operator $\hat{\beta} = \omega - \Im(\frac{FT(\mathcal{D}h.x_2;\omega)}{FT(h.x_2;\omega)})$. This clearly shows that the frequency reassignment and the time reassignment are simultaneous and therefore cannot be dissociated.

The discrete version of the reassignment method can be defined as:

$$\mathcal{H}_r \triangleq \frac{X(t,\omega_k;\mathcal{D}h)}{X(t,\omega_k;h)} \qquad \mathcal{H}_t \triangleq \frac{X(t,\omega_k;\mathcal{T}h)}{X(t,\omega_k;h)}$$

$$\hat{\beta} = \omega_k - \Im(\mathcal{H}_r) \qquad \hat{t} = t + \Re(\mathcal{H}_t)$$

The discrete formulation of the reassignment seems to introduce a small bias in the estimation [21]. In fact, the demonstration presented previously for the continuous FT is not valid anymore for the discrete DFT. This is due to the derivative property of the FT which is only valid for the continuous FT. Fast computation of $X(t,\omega_k;\mathcal{D}h)$ and $X(t,\omega_k;\mathcal{T}h)$ can be derived for particular windows, such as the Hann window, using methods similar to the one used in [17]. For the continuous reassignment, fast approximations can also be derived [21]. In both cases, the complexity can be reduced to one FFT computation.

Algorithm: Frequency reassignment (FR)

1. Compute the FFT for a time t and for the window h, the window $\mathcal{D}h$ and the window $\mathcal{T}h$.
2. For each maximum bin k at the time t
3. Compute $\mathcal{H}_r$, $\mathcal{H}_t$
4. Compute $\hat{\beta} = \omega_k - \Im(\mathcal{H}_r)$. The time of estimation is $\hat{t} = t + \Re(\mathcal{H}_t)$.

5. DISCUSSION ON THE COMMON PRINCIPLES OF THE STFT-BASED FREQUENCY ESTIMATORS

The purpose of this section is to highlight the common principles of the frequency estimators using only STFT time-frequency points.

In order to achieve a direct frequency estimation, we need to eliminate all the other unknown parameters: the amplitude, the initial phase and the slope in the case of the quadratic phase. Let's define $\delta \triangleq \mathrm{e}^{j\gamma\tau_n^2/2}$ and Γ its DFT:

$$\Gamma(\omega,\gamma;h) \triangleq \sum_{n=0}^{N-1} h(\tau_n)\,\mathrm{e}^{j\gamma\frac{\tau_n^2}{2}}\,\mathrm{e}^{-j\omega\tau_n} \qquad (5.1)$$

The STFT of x_2 (3.3) can be put under the form:

$$X(t_m,\omega_k;h) = A\,\mathrm{e}^{j\,\alpha}\,\Gamma(\omega_k-\beta,\gamma;h) \qquad (5.2)$$

As the definition of the quadratic phase model (3.3) is local (i.e. in the neighborhood of the time t_m), α and β depend on t_m. For our problem, we need to express all the STFT points using the parameters α and β for a unique time. The first step is therefore to choose a time-frequency reference (t_0,ω_0). The

STFT of x_2 can be written using the parameters of the time reference. Let's note $\Delta\omega \triangleq \omega_k - \omega_0$ and $\Delta t \triangleq t_m - t_0$.

$$\begin{aligned} X(t_m, \omega_k; h) =& A.\, \mathrm{e}^{j(\alpha+\beta\Delta t+\gamma\frac{\Delta t^2}{2})} \\ & .\Gamma(\omega_k - \beta + \Delta\omega - \Delta t\gamma, \gamma; h) \end{aligned} \quad (5.3)$$

Each STFT point has its own $\Delta\omega$ and Δt values.

Elimination of α and A

This previous equation clearly shows that the amplitude and the initial phase form a complex factor for all the STFT points. From equation (5.3), $\mathrm{e}^{j\alpha}$ can be factorized and the STFT point X_i can be expressed as $X_i = \mathrm{e}^{j\alpha} f_i(A, \beta, \gamma)$, where f_i is a function independent of α. The initial phase can be eliminated in two ways: by division of two STFT points (5.4), or by multiplying one STFT point by the conjugate of a second one (5.5).

$$\begin{aligned} X_2/X_1 &= f_2(A, \beta, \gamma)/f_1(A, \beta, \gamma) \quad (5.4) \\ X_2.X_1^* &= f_2(A, \beta, \gamma).f_1(A, \beta, \gamma)^* \quad (5.5) \end{aligned}$$

In a similar manner, A can be factorized and the STFT point X_i can be expressed as $X_i = A.g_i(\alpha, \beta, \gamma)$, where g_i is a function independent of A. The amplitude can also be eliminated in two ways: by division of two STFT points (5.6) or by taking the argument of an STFT (5.7).

$$\begin{aligned} X_2/X_1 &= g_2(\alpha, \beta, \gamma)/g_1(\alpha, \beta, \gamma) \quad (5.6) \\ \arg(X_1) &= \arg(g_1(\alpha, \beta, \gamma)) \quad (5.7) \end{aligned}$$

All the STFT-based frequency estimation methods should combine these possibilities in order to eliminate both the amplitude and the initial phase.

A linear combination of STFT points preserves this complex factor and can also be used to eliminate the initial phase and the amplitude. This is what is done for a wide number of methods including the phase-vocoder-based methods, the reassignment, the derivative method, some of the DFS methods using phase such as the F-method. All these methods use a ratio of the form:

$$\mathcal{H} = \frac{\sum_i \mu_i X_i}{\sum_i \nu_i X_i} \quad (5.8)$$

where μ_i and ν_i are complex factors which weight the STFT points. The amplitude and the phase are both eliminated through the ratio in this case. When considering the linear phase model, the only unknown parameter left is β. So the previous ratio can be expressed as a function l depending on β only. If this function can be found, and if it is invertible, an estimation of β can be done using l^{-1}: $\hat{\beta} = l^{-1}(\mathcal{H})$.

Macleod's estimator and the parabolic estimator eliminate A and α differently: first the phase is eliminated using the conjugate method. It is obvious in Macleod's estimator, where the function $R_2 = \Re(X_1.X_2^*)$ is used. For the parabolic estimator the modulus is used, but the modulus can be understood as a conjugate product: $|X| = \sqrt{X.X^*}$. A linear combination of these elements R_1 and R_2 preserves the factorization of A, which is therefore eliminated by a ratio of these elements. When considering the linear phase model, the same remark as in the previous paragraph holds.

Elimination of γ

In the case of the quadratic phase model, the problem is more complex because slope and frequency are intimately linked. Indeed a slope multiplied by a time is a particular frequency, so the problem is more to express our equation as a function of a unique frequency $\hat{\beta}$, estimated for a particular time $\hat{t}$.

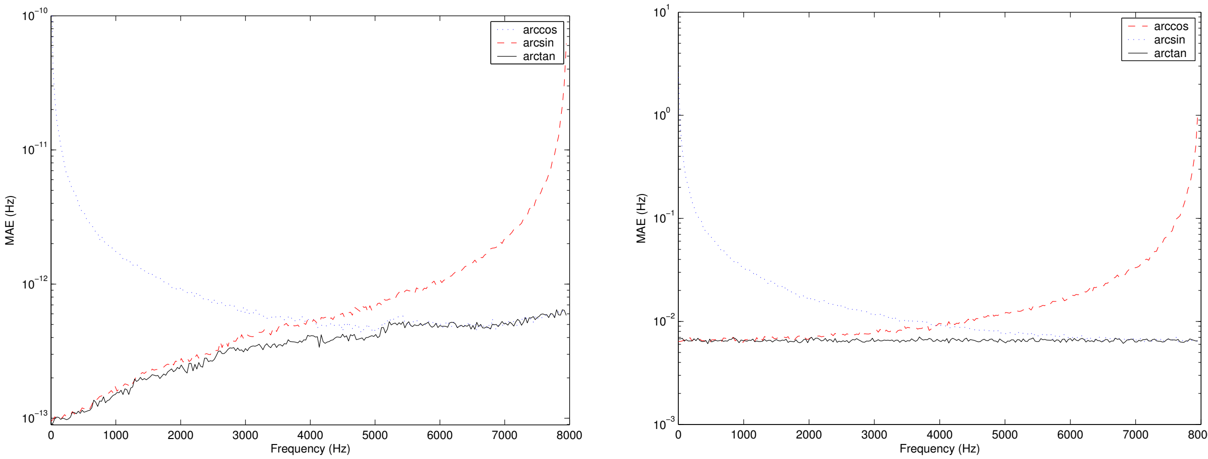
Conclusion

Two problems arise: which combination of parameters in $\mathcal{H}$ can lead to this kind of relation, and how to derive the function $l(\beta)$ from a particular $\mathcal{H}$. The first question is difficult to answer. All the possible ratios $\mathcal{H}$ will not lead to an invertible l function. The second question looks easier. If $\mathcal{H}$ leads to an invertible function of β, an analytical expression of l^{-1} is not absolutely necessary, as it can be modeled by an appropriate function.

6. EVALUATION

Many studies comparing the classical frequency estimation methods in the linear phase case have been done [11], but not for the quadratic phase case, which is another motivation for a broad performance comparison.

In order to achieve a frequency estimation, peak detection is needed, but as our purpose is to compare the frequency estimators, it will be assumed



(a) No noise added

(b) SNR = 40db

Fig. 2: Comparison of the arcsin, arccos and arctan estimators

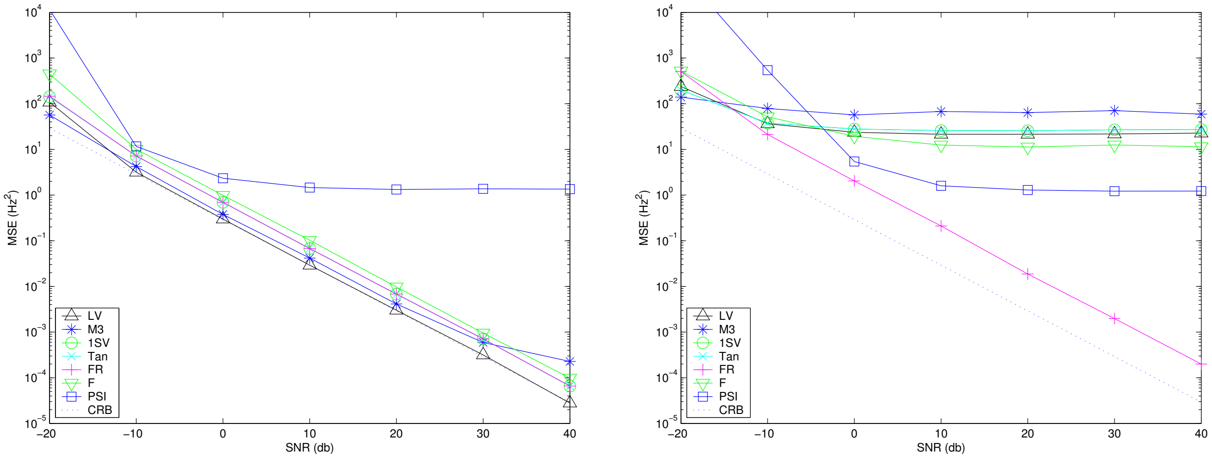
in all experiments that the correct maximum bins are known. The second maximum bin, if needed, is still supposed unknown.

The Cramer-Rao Bound [8] (CRB) for the linear phase model is represented by a dashed line for $N = 512$. It is given here just as an indication for two reasons. First, all the methods presented here do not have the same CRB: the vocoder has a lower CRB because it uses more signal samples for its estimation (two consecutive DFTs separated by H samples correspond to a total of $N + H$ samples). Secondly, for the chirp analysis, the CRB will be different because the parameter γ is added.

The experiments are presented for $F = 16000$ and $N = 512$. All the estimators studied are independent of the initial phase and of the amplitude. The error between the true and estimated values is based on an average of 1000 experiences, using random frequencies inside $[0, 8000]$, and a slope inside $[0, 8000]$ for figure 3(b). For the arcsin, arccos, arctan and 1SV estimator, the hop size is one sample. For the long vocoder (LV), the hop size chosen is 256. For all the methods except M3, the window used is the Hann window.

Figure 2 shows a comparison of the estimators arcsin, arccos and arctan presented in section 4.2. It shows the Mean Absolute Error (MAE) of each estimator as a function of the frequency. On figure 2(a), no noise has been added. All the estimators are unbiased but, for high frequencies (resp. the low frequencies), the variance increases for the arcsin method (resp. the arccos method). These variances values are small on figure 2(a), but when noise is added, the variance grows quickly, as it is shown on figure 2(b) for SNR=40db. The arctan estimator does not have the instability of the other two, and its variance stays low for all the frequencies. In the next experiments, only the arctan estimator is retained.

Figure 3(a) represents the classical performances of the algorithms when the frequency is constant (linear phase). For very low SNRs (<-15db), the performances degrade even if the detection is assumed. For high SNRs, the bias of the algorithms is revealed as the influence of noise becomes negligible. The performance curves of the reassignment, the 1SV and the arctan method are very close. The best performances are obtained with the long vocoder (LV) method, because of the important hop size chosen. In section 4, it has been said that the discrete version of the Reassignment method (FR) presents a bias. This bias begins to appear at 80db and is not visible on this figure. The vocoder, and the arctan method are the only unbiased methods. When the bias influence becomes negligible compared to the noise influence, the performances curves become parallel to the CRB. The shift from the CRB is influenced by the windowing. Intuitively windowing has a tendency of ‘decreasing’ the number of samples used [12], and the corresponding CRB will increase. This fact ap-



(a) In the linear phase case ($\gamma = 0$)

(b) In the quadratic phase case ($0 \leq \gamma \leq 8000$)

Fig. 3: Comparison of estimators as a function of the SNR

pears in the experiment 3(a): Macleod's estimator, which uses a rectangular function, has better performances than all the other estimators which use the Hann window. The main drawback of using the rectangular window appears when multiple sinusoid components are present [12]. The poor performances of the parabolic interpolation are due to the fact that no zero-padding is done, because it has been chosen to compare algorithms with a same order of complexity. Zero-padding consists in using a DFT with a greater length than the signal length (the signal is completed with zeros). It allows to interpolate the amplitude spectrum and is widely used with DFS interpolation methods. For a zero-padding factor of 4, the performances of the parabolic interpolation become comparable to the other methods [7].

A second experiment based on the quadratic phase model has been done. It has been said in the section 3 that for methods based on the linear phase model there is no particular time of estimation. For this experiment it has been assumed that these methods estimate the frequency for the middle of the window. For high slope values (figure 3(b)), the linear phase model is not valid anymore, even approximately, and all the estimators based on this model present unsurprisingly an important bias. The small bias of the discrete version of the reassignment method is unchanged but the variance of the reassignment has increased notably. As the slope grows, the ratio used in Macleod's estimator tends to zero. Therefore, when the slope is high, the estimation tends to the maximum bin estimation. But the maximum bin is supposed to be known for these experiments, which explains why the estimation curve is constant for this estimator. In this experiment, the vocoder and the arctan methods have once again very close performances.

7. CONCLUSION

This work has presented a clear review and description for a wide number of STFT-based frequency estimators. An original demonstration of the reassignment method for the quadratic phase model has been presented, as well as a new and clear justification of the derivative and arccos method for a linear phase model. A new estimator has been presented, the arctan estimator, which belongs to the same family of estimators as the phase vocoder and the derivative method. This estimator avoids the instability of the derivative and the arccos method, and presents very similar results compared to the phase vocoder estimator. An experimental comparison of the estimator described has been performed, with a clear advantage for the phase vocoder method with a long hop size, when no frequency slope is present. The frequency reassignment is the only method which continues to give good results for high frequency slope values. At last, the common principles of all these estimators have been discussed, showing the great potential of exploring new estimators combining differently the STFT points.

8. REFERENCES

[1] R.J. McAulay and T.F. Quatieri, “Speech analysis/synthesis based on a sinusoidal representation,” *IEEE transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 4, pp. 744–754, Aug 1986.

[2] E.G.P. Schuijers, A.W.J. Oomen, A.C. den Brinker, and D.J. Breebaart, “Progress on parametric coding for high quality audio,” *DAGA*, pp. 860–861, Mar 2003.

[3] Xavier Serra, *Musical sound modeling with sinusoids plus noise*, Swets&Zeitlinger, 1997.

[4] Xavier Rodet, “Musical sound signal analysis/synthesis: Sinusoidal+residual and elementary waveform models,” *IEEE Time-Frequency and Time-Scale Workshop*, Aug 1997.

[5] E. Moulines and J. Laroche, “Non-parametric techniques for pitch-scale and time-scale modification of speech,” *Speech communication*, pp. 174–215, 1995.

[6] Myriam Desainte-Catherine and Sylvain Marchand, “High precision Fourier analysis of sounds using signal derivatives,” *J. Audio Eng. Soc.*, vol. 48, no. 7/8, pp. 654, Jul 2000.

[7] Mototsugu Abe and Julius O. Smith III, “Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of FFT magnitude peaks,” *Audio Engineering Society 117th Convention*, 2004.

[8] M. Macleod, “Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones,” *IEEE Transaction on Signal Processing*, vol. 46, no. 1, pp. 141–148, Jan 1998.

[9] Elias Aboutanios and Bernard Mulgrew, “Iterative frequency estimation by interpolation on Fourier coefficients,” *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1237–1241, Apr 2005.

[10] François Auger and Patrick Flandrin, “Improving the readability of time-frequency and time-scale representation by the reassignment method,” *IEEE Transaction on Signal Processing*, vol. 43, no. 5, pp. 1068–1088, May 1995.

[11] Konrad Hofbauer, “Estimating frequency and amplitude of sinusoids in harmonic signals,” Tech. Rep., Graz University of Technology, Apr 2004.

[12] Stephen Hainsworth and Malcolm Macleod, “On sinusoidal parameter estimation,” *Proc. of the 6th Int. Conf. on Digital Audio Effects (DAFx)*, 2003.

[13] Sylvain Marchand, *Modélisation informatique du son musical (analyse, transformation, synthèse)*, Ph.D. thesis, Université de Bordeaux, 2000.

[14] Barry G. Quinn, “Estimating frequency by interpolation using Fourier coefficients,” *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1264–1268, May 1994.

[15] J.L. Flanagan and R.M. Golden, “Phase vocoder,” *Bell System Technical Journal*, pp. 1493–1509, Nov 1966.

[16] Michael R. Portnoff, “Short-time Fourier analysis of sampled speech,” *IEEE transaction on Acoustics, Speech and Signal Processing*, vol. 29, no. 3, pp. 364–374, Jun 1981.

[17] J.C Brown and M.S. Puckette, “A high resolution fundamental frequency determination based on phase changes of the Fourier transform,” *J. Acoust. Soc. Amer.*, vol. 94, pp. 662–667, Aug 1993.

[18] Miller S. Puckette and Judith C. Brown, “Accuracy of frequency estimate using the phase vocoder,” *IEEE Transaction on Speech and Audio Processing*, vol. 6, no. 2, pp. 166–176, Mar 1998.

[19] Mathieu Lagrange, *Modélisation sinusoïdale des sons polyphoniques*, Ph.D. thesis, Université de Bordeaux, 2004.

[20] M. Betser, P. Collen, and G. Richard, “Frequency estimation based on adjacent DFT bins,” *Submitted to EUSIPCO*, 2006.

[21] Stephen Hainsworth and Malcolm Macleod, “Time-frequency reassignment: measures and uses,” *Proc. Cambridge Music Processing Colloquium*, p. 36, 2003.