Estimating partial frequency and frequency slope using reassignment operators

Axel Röbel
IRCAM, Analysis-Synthesis Team, France
email: Axel.Roebel@ircam.fr

Abstract

The estimation of the frequency slope of a partial from its peak in the DFT spectrum today is possible only if a Gaussian window is used. In the following we derive a new method to estimate the frequency slope of a partial from its DFT spectral peak based on the reassignment operators. Compared to the Gaussian window based method our new method can be used with a much larger variety of windows and often achieves better accuracy for equal resolution. After a short introduction into the reassignment method we present a short analytical derivation of the method and we investigate into the analysis properties in relation with the window properties. Based on the analytical derivation of the method we explain the basic requirements for the windows to be used to achieve high accuracy estimates for frequency and frequency slope.

1 Introduction

Analysis of spectral peaks to obtain partial parameters is widely used. One of the main applications is during the analysis stage of additive synthesis models where the partial parameters are estimated from a short time Fourier transform of the signal to be modeled. The standard analysis procedure is based on a quadratic interpolation around the center of a spectral peak in the STFT from which amplitude and frequency at the center of the current frame can be derived (Serra 1997b). In later processing stages the partials obtained for the different frames of the signal have to be connected, which usually is performed by means of neighborhood relations such that connected partials are close in frequency and in amplitude. Other applications include estimation of the fundamental frequency of an harmonic signal (Doval and Rodet 1991). Frequency slope estimation may be used to improve the correct phase adaptation during transformation with the phase vocoder (Bristow-Johnson 2001), even if the standard procedure of the frequency estimation is quite different in this case (Serra 1997a).

Ignoring frequency slope during frequency estimation works well for partials with fixed frequency. However, for partials with moving frequency the frequency estimate becomes poor. The use of the estimate to connect the partials in the neighboring frames in an additive synthesis model may lead to wrong connections if the frequency slope of the partial is unknown. Recently the use of the reassignment method (Auger and Flandrin 1995) has been proposed to improve the frequency estimate for partials with moving frequency (Fitz, Haken, and Christensen 2000), however, only a fixed frequency estimate somewhere inside the analysis window has been considered in this work. During sound transformation in the phase vocoder the phase adjustment will be wrong if the partials frequency slope is not taken into account. In our experiments this results in an additional roughness of the transformed sound after time stretching chirps. A further improvement of all the procedures listed above would be possible if a reliable estimate of the partials linear frequency trajectory in the current frame could be derived.

Currently there exists only a single procedure to derive a frequency slope estimate of a partial from its spectral peak which is based on using a Gaussian window for windowing the signal frames (Marques and Almeida 1986). Assuming that cutting the Gaussian well outside its standard deviation does not change the spectral peak a analytic expression can be derived that allows to derive amplitude, frequency and frequency slope of the partial in the center of the frame. Phase estimation can be obtained by means of a least squares procedure. The estimation procedure has been used successfully to improve phase vocoder transformation (Bristow-Johnson 2001). A drawback of this method, however, is the necessity to use a Gaussian window.

In the following article we will derive a new estimation procedure that may be used to estimate the partial frequency and frequency slope without relying on a Gaussian window. The outline of the article is as follows. In section **2** we review the mathematical formulation of the reassignment method. In section **3** we derive the mathematical equations allowing to estimate partial frequency slopes using the reassignment operators. In section **4** we evaluate the accuracy of the estimates of the new method and compare it with the known estimation procedure based on Gaussian windows. During the evaluation we derive a clear understanding of the window properties that are required to achieve high accuracy frequency and frequency slope estimation with the new method.

2 Reassignment method reviewed

The basic idea of the reassignment method has been developped in the context of time frequency distributions in (Auger and Flandrin 1995). In this investigation it has been shown that by means of reassigning the results obtained by means of a STFT perfectly localized chirps and impulses can be obtained.

In the current context we are interested in analysing a single frame of a discrete Fourier transform. Therefore, compared to the investigation of Auger and Flandrin, we base our analysis on a slightly different formulation of the Fourier transform following

$$M_h(x,w,t)e^{j\Phi_h(x,w,t)} = X_h(w,t) = \int x(u)h(u-t)e^{-jw(u-t)}du, \qquad (1)$$

where $x(t)$ is the signal under investigation $h(t)$ is the analysis window with starting position at $t = 0$ and $X_h(w,t)$ is the Fourier transform of the signal at window position t and frequency w using window h. As shown in eq. (1) the Fourier transform can be expressed by means of its magnitude $M_h(x,w,t)$ and phase $\Phi_h(x,w,t)$.

Based on the Fourier transform specified as in eq. (1) we adapt the results of Auger and Flandrin and obtain for the reassignment operators

$$\begin{aligned} \hat{t}(t,w) &= t - \frac{\partial}{\partial w}\Phi_h(x,t,w) \\ \hat{w}(t,w) &= \frac{\partial}{\partial t}\Phi_h(x,t,w). \end{aligned} \qquad (2)$$

The operators $\hat{t}$ and $\hat{w}$ specify the center of gravity of the signals energy distribution around the time frequency position t,w. It can be shown that for a chirp signal the reassignment operators point exactly onto the frequency trajectory of the signal somewhere in the vicinity of t,w the exact position depending on the Wigner-Ville distribution of the window h.

Making use of the fact that the phase Φ_h of the DFT can be obtained by means of

$$\Phi_h(x,t,w) = \mathrm{imag}\{\log(X_h(w,t))\} \qquad (3)$$

it is easy to verify that the reassignment operators in eq. (2) can be calculated by means of Fourier transforming the signal with to additional windows h_D and h_T as follows

$$\begin{aligned} \hat{t}(t,w) &= t - \mathrm{imag}\{\frac{X_{h_T}(w,t)}{X_h(w,t)}\} \\ \hat{w}(t,w) &= w - \mathrm{imag}\{\frac{X_{h_D}(w,t)}{X_h(w,t)}\}. \end{aligned} \qquad (4)$$

The two windows to be used are

$$\begin{aligned} h_T(t) &= th(t) \\ h_D(t) &= \frac{\partial h(t)}{\partial t}. \end{aligned} \qquad (5)$$

Following the interpretation of the reassignment operators as the location of the center of gravity around the time frequency location t,w they are defined for every point t,w of the Fourier transformation having non zero energy. They can be used to derive a frequency estimate of a chirp in the vicinity of t,w as long as there is only one, or at least one dominating chirp, in the neighborhood of t,w. Using eq. (2) and eq. (5) the notion of neighborhood can be interpreted by means of the central peak of the Fourier transform of the analysis windows $h_T(t)$ and $h_D(t)$.

3 Frequency slope estimation

Making use of the reassignment operators described in the previous section the estimation of the frequency slope of a chirp is conceptually simple. We just need to calculate the derivative of the reassignment operators with respect to the window position t and can directly obtain the frequency slope w' of a partial close to time frequency position t,w by means of

$$w'(t,w) = \frac{\frac{\partial \hat{w}(t,w)}{\partial t}}{\frac{\partial \hat{t}(t,w)}{\partial t}} \qquad (6)$$

Calculating the derivatives with respect to time of eq. (2) using eq. (3) and eq. (4) is straightforward and leads to the following results:

$$\begin{aligned} \frac{\partial \hat{t}(t,w)}{\partial t} &= -\mathrm{real}\{\frac{X_{h_{DT}}(w,t)}{X_h(w,t)}\} \\ &\quad + \mathrm{real}\{\frac{X_{h_D}(w,t)X_{h_T}(w,t)}{X_h(w,t)^2}\} \\ \frac{\partial \hat{w}(t,w)}{\partial t} &= \mathrm{imag}\{\frac{X_{h_{DD}}(w,t)}{X_h(w,t)}\} \\ &\quad - \mathrm{imag}\{(\frac{X_{h_D}(w,t)}{X_h(w,t)})^2\}, \end{aligned} \qquad (7)$$

where again the derivative can be calculated by means of a Fourier transform of the signal using the windows given in eq. (5) and the two further windows

$$\begin{aligned} h_{TD}(t) &= t\frac{\partial h(t)}{\partial t} \\ h_{DD}(t) &= \frac{\partial^2 h(t)}{\partial t^2}. \end{aligned} \qquad (8)$$

Because our objective is the analysis of spectral peaks, only, we will replace the Fourier transform by means of a DFT which will be evaluated at the peak positions, only. Due to the fact that all the windows are nearly band limited the error due to aliasing and circular convolution of the window and signal spectra instead of linear convolution will be negligible.

4 Evaluation

To verify the validity of the results and to assess the errors in a practical application we have investigated the squared error of the estimated frequency trajectories for a number of linear chirps and a non linear chirps. Due to space restrictions we will present our results concerned with linear chirps, only.

Because new estimation procedure requires the analysis window to be differentiated with respect to time twice its side lobes are considerably enforced. Therefore, for high accuracy and robust frequency slope estimation much stronger side lobe rejection is required than for standard applications. Note that this holds true for the standard reassignment operators in eq. (2), too, however, less severe because the window is differentiated with respect to time only once.

Besides using the standard hanning, and blackman window we improve side lobe rejection by means of using a discrete version of a squared hanning window, which is equivalent to a 3-term Nutall window with 30dB side lobe decay:

$$h(t) = (0.5 - 0.5\cos(2\pi t/T))^2 \qquad (9)$$

denoted as *han2* window and a window consisting of a multiplication of hamming and hanning window

$$\begin{aligned} h(t) &= (0.5 - 0.5\cos(2\pi t/T)) \\ &\quad (0.54 - 0.46\cos(2\pi t/T)). \qquad (10) \end{aligned}$$

which will be denoted *hamhan* in the following.

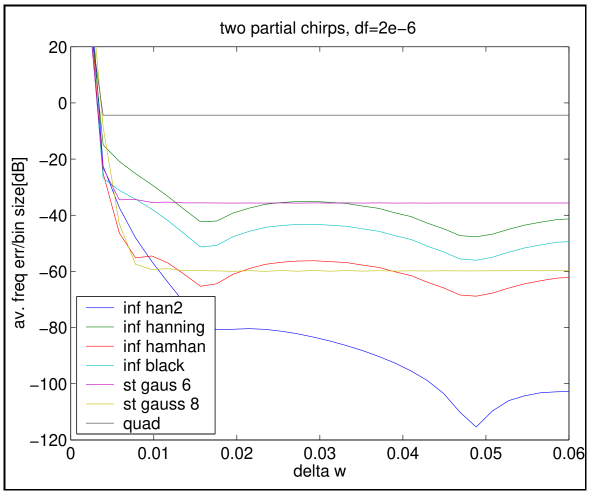

Figure 1: Comparison of frequency trajectory error for different estimation procedures as a function of the analysis window and the frequency distance between two chirps each with frequency slope $dw = 2\cdot 10^{-6}\mathrm{rad/sample}$.

In all cases we compare the results obtained with the reassignment method with the one obtained by means of the Gaussian window as described in (Bristow-Johnson 2001). In case of the Gaussian windows we set sigma to the window length T divided by 6 and 8 respectively which are denoted Gauss6 and Gauss8 respectively. Moreover the result for a simple quadratic interpolation of the log magnitude spectrum and a stationary frequency trajectory is given.

The results of the first experiment are shown in fig. 1. The signal consists of two linear chirps with frequency slope 2e-6 and a frequency offset indicated at the x-axis. The squared average error is indicated in dB relative to the width of a single bin of the 1024 DFT. The window length is 1024. Obviously, all methods that take the frequency slope into account are much better than the stationary model estimating only the center frequency.

All windows used with the new estimation method achieve good results. Due to the fact that the side lobe rejection of the differentiated window is smallest for the hanning and blackman window these two windows obtain less accurate results similar to that obtained with the Gauss6 window. The hamhan and han2 window having much better side lobe rejection for the differentiated windows obtain accuracy and frequency resolution cmoparable to and better than the Gauss8 window.

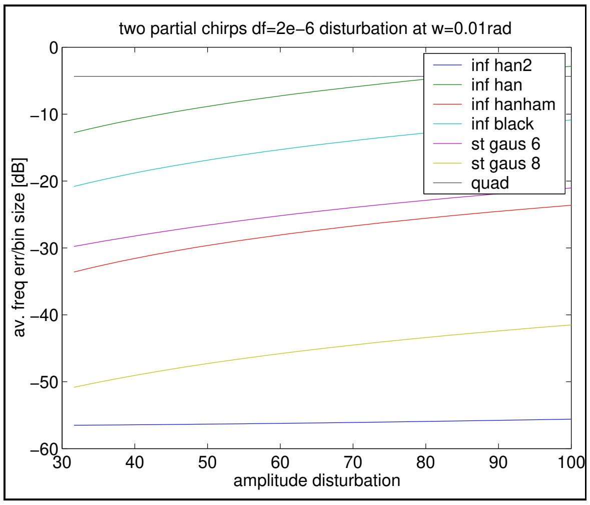

Figure 2: Frequency trajectory error of the different estimation procedures as a function of the analysis window and the amplitude of a disturbing partial at frequency $w = 0.01\mathrm{rad}$.

In the next experiment we investigate the robustness of the frequency slope estimate with respect to dominating partials further away from the peak to be analyzed. Again we estimate the frequency trajectories for two partials starting at frequency $0.06\mathrm{rad}$ and $0.07\mathrm{rad}$ with a frequency slope of $5\cdot 10^{-7}\mathrm{rad/sample}$ and amplitude 1 given a third partial is present at frequency $0.01\mathrm{rad}$ with amplitude a. In figure fig. 2 the frequency trajectory error is depicted as a function of a. Obviously the han2 window is most robust due to its strong attenuation of all side bands for all the analysis windows. The *hanham* window perfoms slightly worse than during the previous experiment, however, is still slightly better then the estimation obtained with the Gauss6 window.

From the results obtained so far it can be concluded that the application of the estimation procedure requires a window with strong side lobe attenuation not only for the window itself but for its first and second time derivative, too. Standard windows as for example hanning or blackman can be applied, however, appear to have only weak robustness against strong peaks far away from the peak to be analysed. The *han2* and *hamhan* window, however, facilitate the use of the method with robustness and accuracy comparable and exceeding the one obtained with the Gaussian windows.

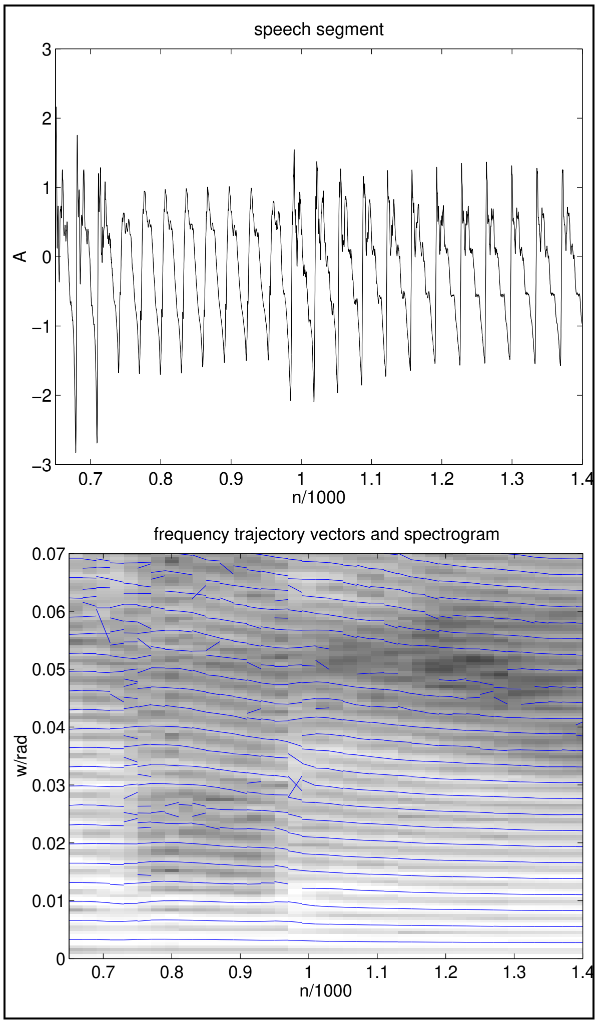

Figure 3: Superposition of frequency vectors estimated for speech segment produce reliably connected trajectories even in case of transients. (note that in the spectrogram dark color represents high amplitude).

4.1 Analyzing a speech signal

As a last example we present the analysis of a real world speech signal. In fig. 3 we have superimposed a small section of the signals spectrogram together with the frequency vectors obtained with the *han2* window. To prevent the analysis of some spurious maxima that are related to side lobes we have filtered all frequency trajectory vectors have their center frequency more than the width of the main lobe of the window away from the maximum. More sophisticated filtering is possible relying on the fact large frequency slopes are quite rare. However, the simple procedure detecting side lobes is sufficient to produce a very clear image of the frequency trajectories of the speech signal.

5 Conclusion and further work

In the current article we have outlined a new estimation procedure for the frequency and frequency slope of partials. The method is based on differentiating the reassignment operators with respect to time and works robust and accurate for windows that provide sufficient side lobe attenuation in the spectra of the window and its time derivatives. The experimental investigation has shown that frequency resolution and accuracy are comparable to or better than the ones obtained from the method based on Gaussian windows. Currently the proposed windows that achieve best results are a squared hanning window or the product of a hanning and a hamming window. Due to the possibility to express the calculations in terms of discrete Fourier transformation of the signal with specialized windows efficient implementation is possible. For future work we are going to investigate more carefully the optimal window that leads to the best compromise between accuracy, robustness and frequency resolution of the result.

References

Auger, F. and P. Flandrin (1995). Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. on Signal Processing 43*(5), 1068–1089.

Bristow-Johnson, R. (2001). Intraframe time-scaling of nonstationary sinusoids within the phase vocoder. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

Doval, B. and X. Rodet (1991). Estimation of fundamental frequency of musical sound signals. In *ICASSP*, pp. 3657–3660 (Vol. V).

Fitz, K., L. Haken, and P. Christensen (2000). Transient preservation under transformation in an additive sound model. In *Proceedings of the International Computer Music Conference, ICMC'2000*, pp. pp.

Marques, J. S. and L. B. Almeida (1986). A background for sinusoid based representation of voiced speech. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 1233–1236.

Serra, M.-H. (1997a). *Musical signal processing*, Chapter Introducing the phase vocoder, pp. 31–91. Studies on New Music Research. Swets & Zeitlinger B. V.

Serra, X. (1997b). *Musical signal processing*, Chapter Musical Sound Modeling with Sinusoids and Noise, pp. 91–122. Studies on New Music Research. Swets & Zeitlinger B. V.