DETECTION OF NON-STATIONARY SINUSOIDS BY USING JOINT FREQUENCY REASSIGNMENT AND NULL-TO-NULL BANDWIDTH

Miroslav Zivanovic

Dept. Ingeniería Eléctrica y Electrónica, Universidad Pública de Navarra
Campus Arrosadía, 31006 Pamplona, Spain
phone: + 34 948169024, fax: + 34 948169720, email: miro@unavarra.es
http://www.unavarra.es/

ABSTRACT

We present a technique that efficiently describes non-stationarity by parametrizing spectral peaks in terms of the frequency reassignment operator variation and null-to-null peak bandwidth. This parameterization provides for a strong separation between the sinusoidal and noise peak classes in a 2D decomposition space. Consequently, the percentage of misclassified peaks is drastically reduced, which significantly simplifies the analysis/synthesis of audio signals in the following processing stages.

1. INTRODUCTION

Representation and identification of non-stationary sinusoidal components is an issue that has frequently been addressed in audio signal processing. Most methods take the STFT as a basis and then look for a convenient way to overcome the inherent stationarity of the spectrogram representation.

Efforts were made in designing, expanding and improving the sinusoidal models by defining a set of parameters that coherently reflect non-stationarity [1, 2]. The key issue was to obtain the correct estimation for those parameters from all the spectral peaks and then to perform a peak selection. Those methods handle well the non-stationarity in the sinusoidal components but their performance significantly degrades for low signal levels or a mismatch to the signal model. Other efforts were directed towards adapting the stationary spectral estimation methods to estimate non-stationary spectra [3, 4]. However, the principal shortcoming of this approach is a very high computational complexity. The reassignment principle [5] improves the energy localization by sharpening out the non-stationary spectra, but the presence of broad-band noise produces energy clustering in the regions where there are no sinusoidal components.

In this paper we present two criteria that aim to properly treat the non-stationary sinusoids. The first criterion provides for the information about a spectral peak's shape by evaluating the variation of the frequency reassignment operator along that peak. By establishing a relationship between the modulation and peak's bandwidth we obtain a compact way to deal with non-stationary sinusoids. The second criterion aims to refine the first one by reducing the number of noise peaks considered sinusoids. This criterion shows that an expected value of the null-to-null bandwidth for the noise peaks is always below that for the peaks representing noise-free sinusoids. A proper combination of the proposed criteria allows us to establish a simple and computationally efficient detection of sinusoidal peaks from the spectrum.

2. THE FREQUENCY REASSIGNMENT VARIATION CRITERION

Let us recall [5] that for each point in the spectrogram of a signal, the time-frequency reassignment vector $r(t, f)$ is defined as:

$$r(t,f) = (t_r, f_r)\ ,$$
$$t_r = -\Re\left\{\frac{X_{th}(t,f)}{X_h(t,f)}\right\}, \qquad f_r = \frac{1}{2\pi}\Im\left\{\frac{X_{dh}(t,f)}{X_h(t,f)}\right\}, \tag{1}$$

where X_h, X_{th} and X_{dh} are the signal's STFTs corresponding to the analysis window h, the ramped window th and the window's derivative dh respectively.

Among various interpretations of $r(t, f)$ the one described in [6] has inspired the design of the herein proposed frequency reassignment variation criterion. Briefly, it states that for the Gaussian window Gh of unit variance, the reassignment vector can be considered gradient to the logarithmic modulus of the STFT:

$$r(t,f) = \nabla \log\left|X_{Gh}(t,f)\right|. \tag{2}$$

For an arbitrary analysis window $r(t, f)$ still follows the steepest descent direction modified by a non-analyticity factor. The expression (2) suggests the existence of an important link between the modulus of the spectrum and its reassignment vector. Indeed, the reassignment vectors show the direction of maxima in the STFT modulus. Consequently, the frequency reassignment operator f_r conveys the information about the shape of a spectral peak at some frequency and time instant. As the nature of a spectral peak (sine/noise) is often related to the peak shape, then we may expect to get a deeper insight into the spectral peak by evaluating f_r for all the frequencies contained in the peak for a given time instant.

To achieve this goal, we introduce the frequency reassignment variation criterion (FRV) for a spectral peak observed at the time instant t_o in the following way:

$$\mathrm{FRV} = \left| \overline{f f_r} - \overline{f}\, \overline{f_r} \right| . \qquad (3)$$

In (3) all the averages are calculated with respect to the normalized peak power spectrum:

$$\frac{\left| X_h(t_o, f) \right|^2}{\int_{\Delta f} \left| X_h(t_o, f) \right|^2 df} . \qquad (4)$$

The integration parameter Δf is the peak width given as a frequency distance between two neighbouring local minima which delimit the given peak in the spectrum.

Let us consider the Gaussian analysis window applied to linear chirps with the chirp rate β and examine the behaviour of the FRV criterion:

$$x(t) = \frac{2^{1/4}}{\sqrt{\lambda}} e^{-\pi\left(\frac{t}{\lambda}\right)^2 + j\pi\beta t^2} \quad , \quad \lambda = \sigma\sqrt{2\pi} \ . \qquad (5)$$

If the analysis window is centred at t_o, the frequency reassignment operator f_r for the signal (5) was shown to be a linear function of frequency [6]:

$$f_r = -\frac{4\pi\lambda}{1+\lambda^4\beta^2} f + \frac{4\pi\lambda\beta\, t_o}{1+\lambda^4\beta^2} \ . \qquad (6)$$

We observe that the peak centre ($f_r = 0$) corresponds to the instantaneous frequency $f = \beta\, t_o$. By combining (3) with (6) and integrating in the interval $[\beta\, t_o - \Delta f/2, \beta\, t_o + \Delta f/2]$ we obtain the following expression for the FRV:

$$\mathrm{FRV} = \frac{4\pi\lambda}{1+\lambda^4\beta^2} \overline{f^2} \ . \qquad (7)$$

The average square frequency in (7) is the signal's mean square bandwidth, which can be expressed after some calculation as:

$$\overline{f^2} = 2\pi\lambda\left(1+\lambda^4\beta^2\right) . \qquad (8)$$

By substituting (8) in (7) we obtain

$$\mathrm{FRV} = 8\pi^2\lambda^2 \ . \qquad (9)$$

The same result is easily obtained for stationary sinusoids too. The absence of β in the last expression means that the noise-free chirps and pure tones are described by the FRV using only one value, assuming λ constant. For other modulation laws (e.g. sinusoidal) and type of analysis window, as well as in presence of additive noise, the FRV spreads but it will still provide for a compact description of non-stationary sinusoids. The reason for this is that the slope change in (6), due to the modulation, is always compensated by the opposite change in the mean square bandwidth in (7). As an illustration, Figure 1 shows how spectral peaks belonging to different types of sinusoidal modulation laws achieve very similar FRV values. We expect, however, an important dispersion in the FRV domain for the noise peaks, due to their stochastic nature. This will indeed be shown in the sections 4 and 5.

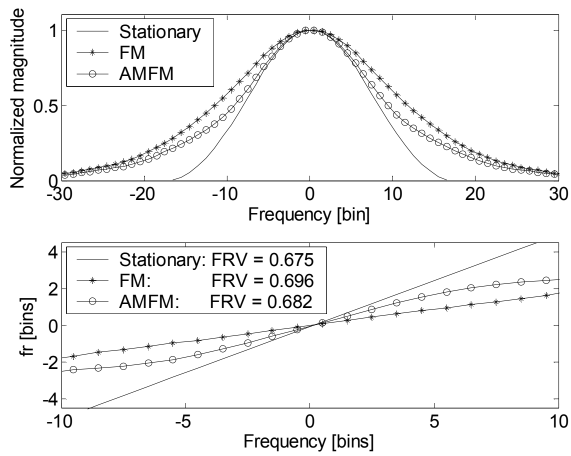

Figure 1 - Illustration of the relationship between the peak shape (above) and the variation rate of the frequency reassignment operator f_r around the peak center (below) for different modulation laws; the analysis window is Hanning.

3. THE NULL-TO-NULL BANDWIDTH CRITERION

The null-to-null bandwidth is usually defined for sinusoidal signals as the spectral mainlobe width of the signal's power spectral density [7]. Although very simple, this parameter is seldom used as an indicator of spectral peak energy concentration, because it does not take into account the shape of the peak. However, combined with some bandwidth-base criterion (like the previously explained FRV) it provides for additional information about a spectral peak and consequently improves its overall description.

Accordingly, for a spectral peak delimited by the frequency range of Δf we define the null-to-null bandwidth criterion (NNB) simply as:

$$\mathrm{NNB} = \Delta f \ . \qquad (10)$$

For a given analysis window the smallest NNB corresponds to the stationary sinusoids while the largest will depend on the type and depth of modulation. Hence, we can not expect (like in the case of the FVR) to achieve a compact description of sinusoidality, because the range of values of the NNB varies proportionally with the maximum modulation in the sinusoidal component. However, we show next that most noise peaks are clustered away from the sinusoidal peaks in the NNB domain i.e. the expected value of NNB for the noise peaks is always smaller than the NNB for stationary sinusoids, independently of the analysis window applied.

Given the random nature of the magnitude spectrum of the Gaussian noise, we must treat the spectral peaks in a probabilistic way. First, let us recall that the magnitude of the spectrum of the Gaussian noise $|X_h(f)|$ is formed by two independent identically distributed Gaussian variables, namely, real and imaginary spectrum. Therefore, the spectrum modulus will have the Rayleigh distribution with the variance dependent on the type of analysis window, while the spectral phase $arg\{X_h(f)\}$ will be uniformly distributed in $(-\pi, \pi)$ [8].

Accordingly, we may consider $|X_h(f)|$ and $arg\{X_h(f)\}$ as the envelope and phase respectively of a wide-sense stationary (WSS) band-pass zero mean Gaussian random process. Next, we make use of an algorithm proposed by Abdi and Nader-Esfahani [9] that estimates the expected number of local maxima per unit time in the envelope $R(t)$ of a wide-sense stationary (WSS) band-pass zero mean Gaussian random process $y(t)$:

$$y(t) = R(t)\cos[2\pi f_c t + \theta(t)] \ . \tag{11}$$

Finally, by using the duality between the time and frequency, we can apply the results from [9] to our problem and thus obtain the estimate for the expected number of maxima per unit frequency in the magnitude noise spectrum $|X_h(f)|$. We shall now give a brief outline of the algorithm which is explained in detail in [9].

Every maximum in $R(t)$ corresponds to a zero with negative slope in $R'(t)$, assuming that $R(t)$ is at least twice differentiable. Hence, the expected number of maxima per unit time N in $R(t)$ is equal to one-half of the expected number of zeros in $R'(t)$ per unit time N_0. According to [9] N can be expressed as follows:

$$N = \frac{1}{2} E[N_0\{R'(t)\}] = \frac{1}{2}\int_{-\infty}^{\infty} |r''| \, p_{R'R''}(0, r'') dr'' \ , \tag{12}$$

where $p_{R'R''}(r', r'')$ is the joint PDF of the random variables $R' = R'(t_o)$ and $R'' = R''(t_o)$ and t_o is an arbitrary time instant. In order to obtain a simple closed-form expression for N, it is common to express (12) in terms of the characteristic function $\Phi_{R'R''}(\xi, \xi')$ of the underlying process R' and its derivative R'':

$$N = -\frac{1}{2\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\xi_2} \frac{d}{d\xi_2} \Phi_{R'R''}(\xi_1, \xi_2) d\xi_1 d\xi_2 \ . \tag{13}$$

The characteristic function $\Phi_{R'R''}(\xi, \xi')$ can be expressed in the following form [9]:

$$\Phi_{R'R''}(\xi_1, \xi_2) = \frac{1}{1 + \alpha_1 \xi_1^2 + \alpha_2 \xi_2^2 - j2B\xi_2^3} \ ,$$

$$\alpha_1 = b_0 b_2 - b_1^2 \ , \qquad \alpha_2 = b_0 b_4 + 3b_2^2 - 4b_1 b_3 \ ,$$

$$B = b_0 b_2 b_4 + 2b_1 b_2 b_3 - b_2^3 - b_0 b_3^2 - b_1^2 b_4 \ . \tag{14}$$

The parameter b_n represents the n[th] central spectral moment of $y(t)$, whose power spectrum $Y(f)$ is centred around the midband frequency f_c:

$$b_n = (2\pi)^n \int_0^{\infty} (f - f_c)^n Y(f) df \ . \tag{15}$$

By introducing (14) in (13) and after some simplifications [9] we obtain the following expression for N:

$$N \approx \frac{1}{2\pi}\sqrt{\frac{\alpha_2}{\alpha_1}} \ . \tag{16}$$

The last result can readily be applied to our problem if we substitute $R(t)$ and $\theta(t)$ in (11) by the Gaussian noise spectrum modulus $|X_h(f)|$ and phase $arg\{X_h(f)\}$ respectively. Then (11) becomes a band-pass stationary Gaussian process whose $Y(f)$, due to the duality of the Fourier transform, will completely be determined by the type of analysis window (as the phase spectrum will always have the same distribution). Consequently, N will be expressed as a number of maxima per unit frequency. Additionally, we can omit f_c in (15) as the envelope of a Gaussian band-pass process is independent of the midband frequency.

Table 1 shows numerical examples for four prominent analysis window types. Their corresponding $Y(f)$ shapes are given in an idealized mathematical form to simplify the calculus of (15). The expected value of the NNB is calculated as a reciprocal of N; the associated values of the NNB for stationary sinusoids are also given in order to facilitate the comparison.

We observe that the expected value of the NNB for the peaks in the spectrum of the Gaussian noise is always smaller than the smallest NNB for the sinusoidal peaks, independently of the type of analysis window. This result is very important because it shows that the NNB criterion has a significant capacity of discerning noise from sinusoids. In the following section we generate a 2D histogram for each peak class and show how the joint action of the FRV and NNB indeed achieves a very efficient sine/noise separation

4. SPECTRAL PEAK HISTOGRAMS

The relationship between the proposed FRV and NNB criterion is established through joint 2D histograms. The criteria were calculated for noise and sinusoidal peaks separately and the corresponding histograms were generated.

Table 1 - Estimated number of maxima per unit frequency N in the magnitude spectrum of Gaussian noise for various analysis window types. NNB(*noise*) = $1/N$ while NNB(*sine*) corresponds to stationary sinusoids.

Window	$Y(f)$	α_1	α_2	N	NNB (*noise*)	NNB (*sine*)
Rectangular	$1,\ 0<f<1$ $0,\ f>1$	3.29	51.95	0.64	1.56	2
Bartlett	$1-\lvert 2f-1\rvert,\ 0<f<1$ $0,\ f>1$	1.64	14.61	0.48	2.08	4
Hanning	$0.5(1-\cos 2\pi f),\ 0<f<1$ $0,\ f>1$	1.28	9.00	0.43	2.32	4
Blackman	$0.42-0.5\cos 2\pi f+0.08\cos 4\pi f,\ 0<f<1$ $0,\ f>1$	1.00	5.67	0.38	2.63	6

In both cases we have first established a signal model, next we have calculated the Hanning windowed STFT and finally we have analyzed the corresponding peaks in the spectrum modulus and generated the corresponding joint histograms. For the noise histograms we have applied a Gaussian zero-mean signal and analyzed all the peaks in the STFT modulus. For the sinusoidal distribution we have taken the following AM-FM vibrato-like signal model:

$$s(n)=\cos\left[2\pi F_0 n + A_{FM}\sin\left(2\pi F_{FM}n+\alpha\right)\right]\times \left[1+A_{AM}\cos\left(2\pi F_{AM}n+\beta\right)\right] . \quad (17)$$

The signal parameters are selected in such a way to ensure that the sinusoidal spectrum always contains a mainlobe i.e. the modulation bands are not resolved. For the size of the analysis window L we have: F_{AM} = 1/(2*L), F_{FM} = F_{AM}/2, A_{AM} = 0.5, A_{FM} = 10; α and β determine the phase relationship between the modulation laws. The corresponding histogram is thus generated by analyzing only the strongest peak in the spectrum of the signal for all the combinations of α and β.

In order to best visualize the decomposition of the peak classes, we have applied a binary mask to each joint histogram by assigning a same value to all the cells in the 2D space belonging to the same histogram. Then the masked histograms were shown together on Figure 2. The conclusions from the previous sections can readily be checked by examining the relative positions of the masked histograms. Where the FRV is considered, all sinusoidal peaks are clustered within a small range while the noise peaks exhibit a significant dispersion. As for the NNB, most sinusoidal peaks are wider than the widest noise peaks. Note that the narrower sinusoidal peaks can still be successfully detected thanks to the particular quasi-triangular shape of the noise masked histogram. The small overlap between the classes corresponds to the segments of $s(n)$ close to stationary sinusoids.

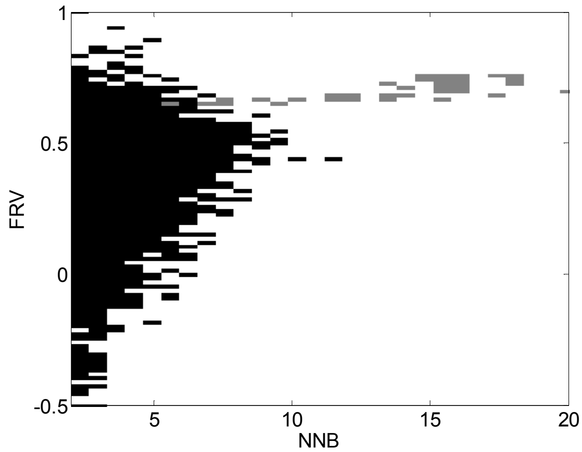

Figure 2 – Joint FRV-NNB masked histograms: sinusoidal peaks (gray), noise peaks (black).

5. EXPERIMENTAL RESULTS

In order to check the validity of the proposed algorithm we have chosen for the comparison the method [1] (from now on the correlation method), which is an important reference in non-stationary sinusoidal modelling. From a spectral peak the correlation method estimates the amplitude and frequency modulation parameters (A, f, Δ_a , Δ_f) which are then used to generate the corresponding signal $a(t)$:

$$a(t)=A\times 10^{\Delta_a t/20}\cos[\phi(t)]$$

$$\phi(t)=\phi(0)+2\pi\int_0^t\left[f+\Delta_f u\right]du . \quad (18)$$

Then, the spectral mainlobe of $a(t)$ is compared to the given spectral peak by means of complex cross-correlation and if the result is significant, the peak is considered sinusoidal.

For the correlation method, we have first generated 1D

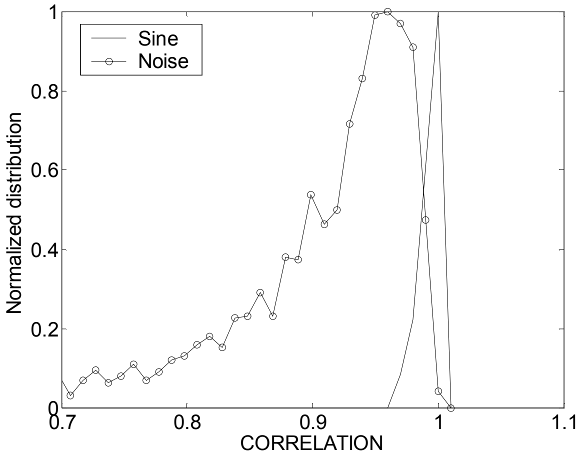

Figure 3 - The correlation method spectral peak histograms.

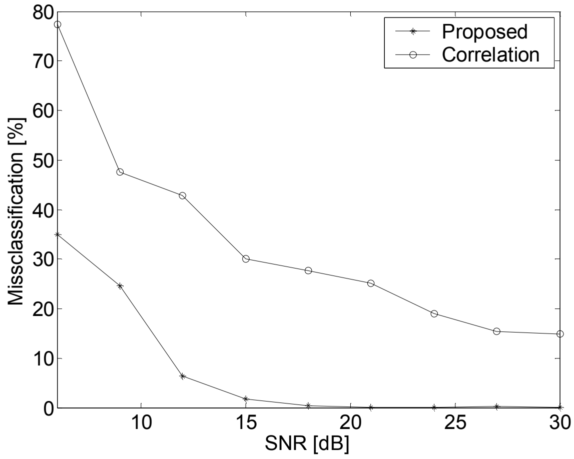

Figure 4 - Percentage of misclassified noise peaks versus local SNR for the correlation and proposed method.

histograms for the sinusoidal and noise peak class by using the same respective models described in Section 4. The corresponding histograms are shown on Figure 3. We observe that although the sinusoidal peaks are all clustered around the maximum correlation equal to one, there is still an important overlap with the noise class. This overlap is basically due to the model mismatch between (17) and (18).

Next, we have investigated into the behaviour of the sinusoidal histograms as the noise in the signal model becomes gradually more significant. As a measure of the influence of the additive noise, we have adopted the local SNR, defined as the decibel ratio between the sinusoidal peak level and the mean noise level in the peak's neighbourhood. Then, we have established the following comparison criterion: for each local SNR we collect all sinusoidal peaks in the spectrogram and we calculate the number of misclassified noise peaks i.e. the noise peaks considered sinusoidal.

Figure 4 shows that for the SNR above 18dB approximately the proposed method removes almost all noise peaks, thanks to the compact description of non-stationarity, as explained in Section 2 and 3. For lower signal levels the sinusoidal histogram from Figure 2 (Section 4) gets more overlapped with the noise histogram. However, we discovered that the sinusoidal histogram exhibits negligible dispersion in the FRV domain, thus preserving its horizontal-oriented shape. Consequently, for the SNR as low as 6dB the sinusoidal histogram overlaps with 34% of the noise peaks opposite to 78% of the noise peaks in the correlation method. Note also that the proposed algorithm is computationally more efficient, as we avoid generating a sinusoidal signal each time the peak correlation has to be performed.

6. CONCLUSIONS

We have presented a compact and simple way to describe non-stationarity in the time-frequency plane. It was shown that the variation of the frequency reassignment operator along the spectral peak, together with the width of the peak, can be joined into a single sinusoidal description. Its low computational complexity and robustness against noise make the proposed method a good candidate for non-stationary sinusoidal detection in audio signals.

7. REFERENCES

[1] M. Lagrange et al, "Sinusoidal Parameter Estimation in a Non-Stationary Model", *Proceedings of the 5th International Conference on Digital Audio Effects*, Hamburg, Germany, September 26-28, 2002, pp. 59-64

[2] M. Abe and J. Smith, "AM/FM Rate Estimation for Time-Varying Sinusoidal Modeling", *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, Philadelphia, USA, March 18-23, 2005, Vol. III, pp. 201-204

[3] J. Pitton, "Adapting Multitaper Spectrograms to Local Frequency Modulation", *Proceedings of the IEEE Workshop on Statistical Signal and Array Processing*, Pennsylvania, USA, August 14-16, 2000, pp.108-112

[4] J. Xiao, P. Flandrin, "Multitaper Time-Frequency Reassignment for Nonstationary Spectrum Estimation and Chirp Enhancement", *IEEE Transactions on Signal Processing*, Vol. 55, No. 6, pp. 2851-2860, June 2007

[5] F. Auger, P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method", *IEEE Transactions on Signal Processing*, Vol.43, No. 5, pp. 1068-1089, May 1995

[6] E. Chassande-Mottin, I. Daubechies, "Differential reassignment", *IEEE Signal Processing Letters*, Vol.4, No.10, pp.293-294, October 1997

[7] H. Nguyen, "Performance of multicarrier DS-CDMA systems with time-limited chip waveforms", *Canadian Journal of Electrical and Computer Engineering*, Vol. 29, NO.1/2, January/April 2004

[8] P. Peebles, "Probability, random variables and random signal principles", McGraw-Hill Inc., 1993

[9] A. Abdi, S. Nader-Esfahani, "Expected number of maxima in the envelope of a spherically invariant random process", *IEEE Transactions on Information Theory*, Vol. 49, NO.5, pp. 1369-1375, May 2003