Audio Engineering Society

Convention Paper

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

Harmonic plus noise decomposition: time-frequency reassignment versus a subspace based method

Bertrand David[1], Valentin Emiya[1], Roland Badeau[1] and Yves Grenier[1]

[1]*ENST, dept TSI, 46 rue Barrault, 75634 Paris Cedex 13, France*

Correspondence should be addressed to Bertrand David (bertrand.david@enst.fr)

ABSTRACT

This work deals with the Harmonic+Noise decomposition and, as targeted application, to extract transient background noise surrounded by a signal having a strong harmonic content (speech for instance). In that perspective, a method based on the reassigned spectrum and a High Resolution subspace tracker are compared, both on simulations and in a more realistic manner. The reassignment re-localizes the time-frequency energy around a given pair (analysis time index, analysis frequency bin) while the High Resolution method benefits from a characterization of the signal in terms of a space spanned by the harmonic content and a space spanned by the stochastic content. Both methods are adaptive and the estimations are updated from a sample to the next.

1. INTRODUCTION

In the context of musical signal processing [1], or audio coding (*cf.* MPEG4-HILN coder), or in the case of some specific forensic application where extracting weak audio transients buried in a sinusoidal foreground [2] is intended, one may need to efficiently decompose the signal into a sinusoidal part (also denominated as the harmonic part, or the deterministic part) and a noisy part (also denominated as the stochastic part or the residual).

More precisely, the model is that of M slowly varying complex exponentials, hence encompasses the case of real data, summed with a stochastic process [3], written down as:

$$s(t) = \sum_{k=1}^{M} b_k(t) \exp(j\Phi_k(t)) + w(t), \qquad (1)$$

where $t \in \mathbb{Z}$ denotes the discrete time index, M the order of the model —*e.g.* the number of complex exponentials, being even, $M = 2P$, when data is

composed of P real sinusoids —, $b_k(t) \geq 0$ the modulation law relative to the kth component magnitude (real). $\Phi_k(t)$ is the instantaneous phase of the component and is bound up to its instantaneous frequency $f_k(t)$ by differentiation:

$$\Phi_k(t)' = 2\pi f_k(t). \quad (2)$$

Note that the frequencies are *not* assumed to be multiples of some fundamental. The stochastic process $w(t)$ may describe several kinds of physical signals : background measurement noise, turbulence noise imputable to air friction when dealing with wind instruments or voice, impulse-shaped, transient noise when processing for instance the onset of a piano or a percussion sound.

Estimation of the model. Since the instantaneous amplitudes and frequencies b_k's and f_k's are expected to be varying, both the parameters of the sinusoidal part and the statistical properties of the stochastic process may be considered as *non-stationary*. To overcome this difficulty, most methods (*cf.* [3, 4]) tend to use a sequential estimation technique applied on overlapping segments of finite length along which a definite sinusoidal model is estimated. For instance, the phase is often taken as a polynomial of low degree (*typ.* 1 or 2) in the variable t. The process $w(t)$ is obtained as a residual, by subtracting the estimated deterministic part. A broad bulk of existing algorithms relies on a time-frequency analysis of the signal, facing the challenging trade-off of shortening the segments for more adequation to the assumption of stationarity while loosing frequency resolution and hence, leading to poor estimates.

In this paper, both answers to this issue concerning harmonic plus noise decomposition are compared: one is based on the reassigned spectrum [5] and the second one is an adaptive subspace based analysis. The methods are described separately in the following sections while the results are demonstrated afterward. More specifically, this work focuses on the ability of each method to extract the noise part while preserving its spectro-temporal shape. Clues on frequency estimation performance can be found in [6, 7, 8] and are not in the scope of this work.

2. HARMONIC+NOISE DECOMPOSITION WITH REASSIGNED SPECTRUM

2.1. Principles

Reassignment operators [5] . The derivation of the so-called reassignment operators in time and frequency relies on the continuous time definition of the Short Time Fourier Transform (STFT). Let be $s_a(t),\ t \in \mathbb{R}$, the analyzed signal, the associated STFT is formulated as:

$$\tilde{S}_a(\tau, f) = \int_{t\in\mathbb{R}} s_a(t)h(t-\tau)e^{-j2\pi ft}dt \quad (3)$$

When facing the problem of localizing amplitude and frequency-modulated sinusoids, the performance limitation is mainly due to the window length and its spectral width (of referred to as the time-frequency box). The reassignment tries to overcome this Fourier-related constraint using the STFT phase information. The STFT is now rewritten in terms of magnitude and phase:

$$\tilde{S}_a(\tau, f) = M(\tau, f)e^{j\varphi(\tau,f)}. \quad (4)$$

The reassignment operators are derived from the partial derivatives of $\varphi(t, f)$ with respect to each of its variables, leading respectively to the instantaneous frequency

$$F_i(\tau, f) = \frac{1}{2\pi}\frac{\partial\varphi(\tau, f)}{\partial\tau}, \quad (5)$$

and to the group delay

$$T_g(\tau, f) = -\frac{1}{2\pi}\frac{\partial\varphi(\tau, f)}{\partial f}. \quad (6)$$

These equations are often interpreted as follows. When considering the energy $M(\tau_0, f_0)^2$ spread around a given point (τ_0, f_0) of the time-frequency plane, its centroid is the point of normalized frequency $F_i(\tau_0, f_0)$ and discrete time $\tau_0 + T_g(\tau_0, f_0)$. Each energy coefficient is said to be *reassigned* to this centroid. The time-frequency content of the signal is then re-mapped on the plane.

2.2. Discrete-time implementation

The Short Time Fourier Transform (STFT) of the sampled data sequence $s(t),\ t \in \mathbb{Z}$ is defined as

$$\tilde{S}(\tau, \nu_k) = \sum_{t=\tau}^{\tau+N-1} s(t)h(t-\tau)e^{-j2\pi\nu_k t}, \quad (7)$$

where $\tau \in \mathbb{Z}$ is the analysis time lag, $\nu_k = k/K$ the frequency bin and $h(t)$ the window applied, assumed to be of finite length N. The order K of the transform has to be greater or equal to N, and is chosen as $K = 2N$ in our practical implementations. The STFT is then rewritten in its polar form as

$$\tilde{S}(\tau, \nu_k) = M(\tau, \nu_k) e^{j\varphi(\tau,\nu_k)}. \quad (8)$$

To approximate the continuous variable derivatives needed in equations (5) and (6) in the context of numerical processing, a numerical filter is used. This filter can be for instance designed with the help of a Remez-Parks-McLellan algorithm for linear phase Finite Impulse Response (FIR) filter. In this work, a different technique is employed, the starting point of which is a polynomial fitting of the sequence [9].

$\varphi(\tau, \nu_k)$ is then extracted and unwrapped for each channel k and derivated to obtain the instantaneous frequency $F_i(\tau, k)$. The same procedure is applied along the frequency axis, yielding the group delay $T_g(\tau, k)$.

Adaptive computation. The algorithm is intended to work with a hop size of only one sample ($(N-1)-$samples overlap). To lower the complexity from the well-known $O(N \log(N))$ cost per sample to a linear one ($O(N)$) the STFT derivation is made adaptive [10, 11]. This gain benefits from the fact that a number of common windows are built with sines and thus, can be written as a sum of geometric sequences of the complex exponential form.

Let for instance the window $h(t)$ be the Hann window:

$$h(t) = \frac{1}{2}(1 - \cos(\frac{2\pi}{N} t)). \quad (9)$$

This is rewritten as

$$h(t) = \frac{1}{2}(1 - \frac{1}{2}(W_N^t + W_N^{-t})),\ t \in [0,\ N-1],$$

where $W_N = e^{j2\pi/N}$, which leads to a decomposition of the STFT:

$$\tilde{S}(\tau, \nu_k) = 0.5\tilde{S}_0(\tau, \nu_k) - 0.25(\tilde{S}_1(\tau, \nu_k) + \tilde{S}_2(\tau, \nu_k)), \quad (10)$$

where $\tilde{S}_0(\tau, \nu_k)$ is the STFT using the rectangular window $u_N(t) = 1,\ t \in [0,\ N-1]$ and $u_N(t) = 0$ otherwise, and where $\tilde{S}_1(\tau, \nu_k)$ and $\tilde{S}_2(\tau, \nu_k)$ are the STFT respectively windowed by $W_N^t u_N(t)$ and $W_N^{-t} u_N(t)$.

Defining the simple increment

$$\Delta s(\tau, k) = e^{-j2\pi\nu_k\tau} ((-1)^k s(\tau + N) - x(\tau)), \quad (11)$$

an update of each STFT is readily obtained as

$$\begin{cases} \tilde{S}_0(\tau+1, \nu_k) = \tilde{S}_0(\tau, \nu_k) + \Delta s(\tau, k) \\ \tilde{S}_1(\tau+1, \nu_k) = \tilde{S}_1(\tau, \nu_k) + W_N^{-1} \Delta s(\tau, k) \\ \tilde{S}_2(\tau+1, \nu_k) = \tilde{S}_2(\tau, \nu_k) + W_N \Delta s(\tau, k) \end{cases} \quad (12)$$

The update of the whole STFT then results from the equation (10).

Harmonic+noise decomposition

The reassignment principles have been applied for enhancing the time-frequency representation, for frequency estimation [5, 12] and also for source/filter modeling in speech processing [13]. As the formulae 5 and 6 cited above result in the precise localization of the frequency estimates, a reconstruction technique is to be determined to extract the harmonic part on one side and the noise on the other. As in many other works, the former is obtained at first and subtracted afterward from the original to get the latter.

For a given segment of analyzed data located in the interval $[\tau,\ \tau + N - 1]$, the *Harmonic part* of the signal is computed following the steps:

1. STFT computation and peak-picking of its magnitude,
2. derivation of $F_i(\tau, k)$ and $T_g(\tau, k)$ for each peak k,
3. selection among this collection of peaks of the bins l where the instantaneous frequency and the bin frequency match, *i.e* F_i must lie in the vicinity of the frequency center of the channel, for instance:

$$|F_i(\tau, l) - l/K| < \frac{3}{2}(1/2K). \quad (13)$$

This stage can be post-processed by a median filtering to remove isolated points,

4. for each selected bin l, a complex exponential at the frequency $F_i(\tau, l)$ is computed with an amplitude taking into account the phase and amplitude distorsion due to windowing at the frequency $F_i(\tau, l)$,

5. the synthetized component is added to the output segment, windowed by a Hann window centered on the time-instant $\tau + N/2 + T_g(\tau, l)$.

It is worth making mention here that the Hann window utilized for the synthesis is not of constant length, since it depends on the reallocation time in the analyzed interval. Let $L_h(\tau, l)$ be this length, this is expressed as:

$$L_h(\tau, l) = N - 2|T_g(\tau, l)|. \tag{14}$$

In addition, for approaching perfect reconstruction, the synthesis window $h_s(t)$ is weighted by the factor $(\sum_{t=0}^{L_h-1} h_s(t))^{-1}$ to be made unitary.

Once the steps 1-5 have been repeated all along the analyzed signal, the harmonic part $s_h(t)$ is derived. The noise part is then deducted as:

$$s_n(t) = s(t) - s_h(t) \tag{15}$$

3. ADAPTIVE HIGH RESOLUTION HNM DECOMPOSITION

Since the end of the 18th century [14, 15], Fourier analysis and High Resolution (HR) methods have been both complementary and competitors. While the former developed into the prominent tool in the field of the spectral analysis, the latter has revealed himself in the two last decades to be one of the most valuable estimation technique in the so-called Direction Of Arrival problem [16]. Notwithstanding its remarkable resolution properties, its use remains marginal in audio processing tasks, even though the underlying model is well adapted for tracking slow varying line spectra [17].

3.1. Theoretical background

Subspace analysis. Subspace decomposition is the theoretical foundation of a number of methods (Pisarenko [18], MUSIC [19], Matrix Pencil [7], ESPRIT [20]). The subspace analysis relies on the following remark. Let $x(t)$, $t \in \mathbb{Z}$ be a complex signal, linear combination of M complex exponentials:

$$x(t) = b_0 z_0^t + b_1 z_1^t + \ldots + b_{M-1} z_{M-1}^t, \tag{16}$$

where the z_k's, $k = 0, 1, \ldots, M-1$, are the complex poles of the signal and b_k's the associated complex amplitudes. More precisely, $z_k = \exp(\delta_k + j2\pi\nu_k)$ where $\delta_k \in \mathbb{R}$ is the damping or growing factor and $\nu_k \in [-0.5\ 0.5]$ is the normalized frequency. Expanding this definition to the vector of the n ($n \geq M$) subsequent samples $\mathbf{x} = \begin{bmatrix} x(0) & x(1) & \ldots & x(n-1) \end{bmatrix}^T$ leads to the matrix expression :

$$\mathbf{x} = \mathbf{V}\mathbf{b}, \tag{17}$$

where $\mathbf{b} = \begin{bmatrix} b_0 & b_1 & \ldots & b_{M-1} \end{bmatrix}^T$ and $\mathbf{V}$ is the Vandermonde matrix defined as:

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{M-1} \\ z_0^2 & z_1^2 & \ldots & z_{M-1}^2 \\ \vdots & \vdots & \vdots & \vdots \\ z_0^{n-1} & z_1^{n-1} & \ldots & z_{M-1}^{n-1} \end{bmatrix} \tag{18}$$

For M distinct poles, the M vectors $\{\mathbf{v}(z_k)\}_{k=0,1,\ldots,M-1}$, defined as the column vectors of the matrix $\mathbf{V}$, $\mathbf{v}(z_k) = \begin{bmatrix} 1 & z_k & \ldots & z_k^{n-1} \end{bmatrix}^T$, are linearly independant. Thus the range space of $\mathbf{V}$ is of dimension M. In short, a vector of n subsequent samples of a signal combining linearly M complex exponential belongs to a M dimensional subspace, the so-called *signal subspace*. When dealing with a noisy signal model : $s(t) = x(t) + w(t)$, the vector $\mathbf{s} = \begin{bmatrix} s(0) & s(1) & \ldots & s(n-1) \end{bmatrix}^T$ belongs to a n-dimensional subspace. Under the hypothesis of a Wide Sense Stationary (WSS) white noise, this subspace can be decomposed as the direct sum of the M-dimensional signal subspace and its orthogonal complementary, of dimension $n - M$, referred to as *the noise subspace.*

Harmonic+noise decomposition. Let $\mathbf{W}$ be a $n \times M$ matrix, conveniently chosen as orthonormal, whose range space is the signal subspace. The projection matrices onto the signal subspace and onto the noise subspace are thus respectively $\mathbf{P}_s = \mathbf{W}\mathbf{W}^H$ and $\mathbf{P}_n = \mathbf{I} - \mathbf{P}_s$, where the subscript H denotes the hermitian transpose. For a given vector of data $\mathbf{s}$, the harmonic part is then obtained by:

$$\mathbf{s}_h = \mathbf{P}_s \mathbf{s} \tag{19}$$

while the noise part is the reminder:

$$\mathbf{s}_n = \mathbf{P}_n \mathbf{s} \tag{20}$$

These expressions need two remarks:

- even in the ideal case of stable signal components (neither amplitude nor frequency modulation) and WSS white noise, this decomposition *does not* lead to $s_h(t) = x(t)$, simply because considering a noise vector of n subsequent samples $\mathbf{w}$, this vector usually belongs to a n-dimensional space in which the noise subspace as defined above is included;
- neither estimation of the parameters (frequencies, damping factors, amplitudes) has to be made explicitly.

Tracking of W. In a number of methods, a matrix $\mathbf{W}$, the columns of which form a basis of the signal subspace, is derived by means of a Singular Value Decomposition of the covariance matrix $\mathbf{C}_{ss}$ of the data. Conversely, the subspace method used in this work is adaptive, referred to as the Fast Approximated Power Iteration in the literature [21]. Starting from a rank one update of the covariance matrix,

$$\mathbf{C}_{ss}(t) = \beta \mathbf{C}_{ss}(t-1) + \mathbf{s}(t)\mathbf{s}(t)^T, \qquad (21)$$

where $\beta < 1$ is a real positive forgetting factor and $\mathbf{s}(t) = \begin{bmatrix} s(t) & s(t+1) & \dots & s(t+n-1) \end{bmatrix}^T$, it conduces in $3nM + O(M^2)$ operations[1] to a rank one update of the form:

$$\mathbf{W}(t) = \mathbf{W}(t-1) + \mathbf{e}(t)\mathbf{g}(\mathbf{t})^H \qquad (22)$$

where $\mathbf{e}(t)$ and $\mathbf{g}(t)$ are column vectors. The whole description of the algorithm is beyond the scope of this paper and can be found in [21].

3.2. Preprocessing

As the method relies on a model comprehending an additive white stationary noise process, its performances lower when dealing with real signals the stochastic part of which is usually not white. In addition of the coloration of the noise, it is not rare in the audio field to encounter large dynamics. The estimation of the weak harmonic components, often settled in the upper part of the spectrum as low as 40 or 60 dB under the maximum, is then made dubious. The preprocessing designed for applying successfully the subspace tracker described above includes 3 steps:

1. pre-emphasis of the entire signal,
2. subband decomposition,
3. whitening in each subband.

Pre-emphasis and whitening. The first and third preprocessing steps are based on the same principle. The Power Spectral Density (PSD) of the considered sequence is estimated (for instance by means of a Welch-averaged periodogram) and an estimator of the noise PSD is derived as the non-linear median filtering of it. The corresponding AR coefficients are computed for a pre-defined model order K. As the aim of the pre-emphasis is a spectrum detrending, a low order K is chosen for the first step. In each subband, on the contrary, K will be of order 10 to 20, for the noise coloration must be drastically reduced. An exemple of the pre-emphasis of voice segment is given in figure 1. The original signal has then been filtered by a Finite Impulse Response (FIR) filter of length 5 to obtain the pre-emphasized signal.

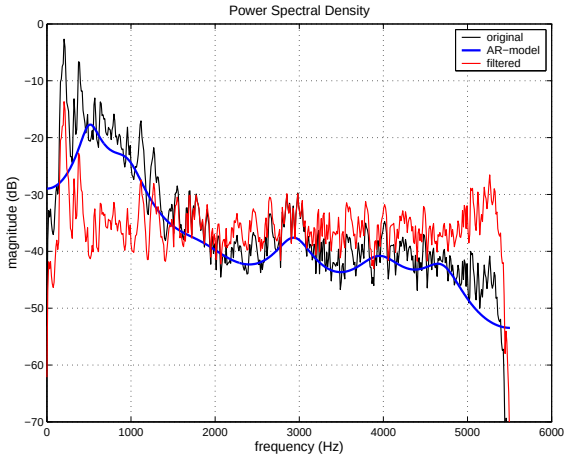

Fig. 1: Pre-emphasis of a voice segment of one second length. The model order is $K = 12$.

Filter bank. The subband decomposition is completed using a quasi-perfect reconstruction cosine-modulated filter bank [22]. Each subband signal is maximally decimated. The adjustment of the subband number depends on the sampling frequency and on the density of harmonic components in the resulting subband. Usual values vary from 4 to 16.

[1] an operation being defined as a Multiply and ACcumulate operation, MAC.

It can be noticed that even if in this work, only uniform filter banks are considered, an extension to non-uniform ones is readily obtained by dyadic iteration. An exemple of uniform 4-subbands decomposition is displayed on the figure 2.

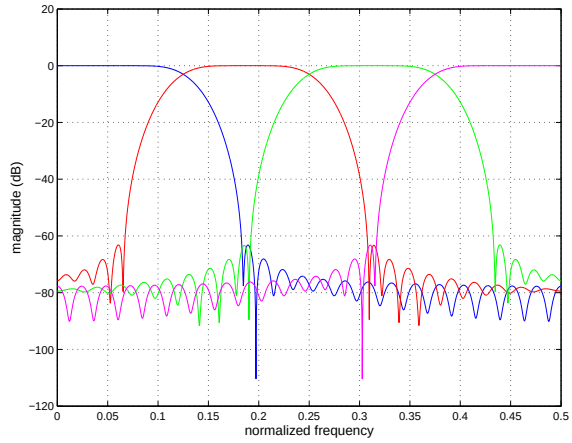

Fig. 2: Cosine modulated analysis filter bank, with 4 subbands.

4. EXPERIMENTS

The aim of this section is to demonstrate the abilities of both algorithms (Reassigned Spectrum-based or Subspace-based) in the task of extracting a background stochastic process and especially a transient (highly non-stationary process) from a signal including a strong harmonic content, speech being the target example of such a kind of signal. To be able to assess the results, a procedure for bringing into existence a non-stationary, impulsive-like process with known characteristics has been defined. The algorithms are then applied to various simulations to give some clues on the parameters tuning according to the context.

4.1. Creating a synthetic non stationary noise

The time-frequency profile of the process is defined as follows.

1. the spectrum at $t = 0$, the initial time instant is defined with the help of a set of poles, leading to an AutoRegressive (AR) spectrum,
2. a low f_l and a high f_h spectral limits are set, and a damping factor $\alpha(f_l)$ is defined for the low limit, owing to which the decreasing of the process around the frequency f_l is of the form $d(t) \propto \exp(-\alpha(f_l)t)$,
3. a damping law is given, as a power function of frequency, *i.e.* $\alpha(f) = \alpha(f_l)(\frac{f}{f_l})^p$

The whole operation is implemented by FFT-filtering of a white stationary noise. An example is drawn on figure 3, obtained at a sampling frequency of 8kHz with the following parameters: a solely pole of 0.99 magnitude at the frequency of 500 Hz, $f_l = 150$ Hz and $f_h = 3500$ Hz, $\alpha(f_l) = 4\,\mathrm{s}^{-1}$ and $p = 1$

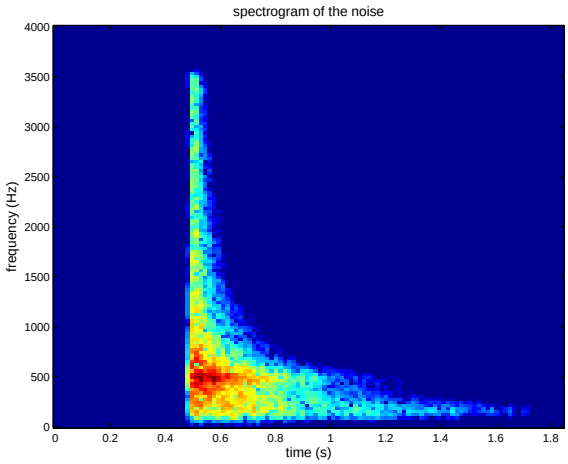

Fig. 3: Time-frequency (256 pts-FFT, Hann windowed) representation of the non-stationary noise.

4.2. Illustrative simulations

All the simulations of this section include a transient noise generated as described above in the section 4.1, and a white background stationary noise around 50 dB below the maximal signal power (this corresponds to an overall Signal To Noise Ratio around -25dB for the whole observation window). In the following, the Fourier-based method is referred to as RF-HND (Reassigned Fourier-Harmonic+Noise Decomposition), while the Subspace analysis-based method is abbreviated as HR-HND.

For each case, the time-frequency representation of the results are given, derived with a Hann windowed 256-points-FFT and jointly scaled to be comparable.

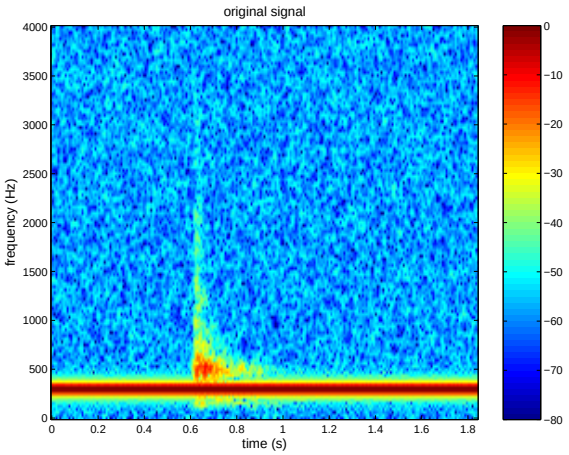

Fig. 4: Time-frequency representation of the original signal.

Pure Sine + noise. In this example, a 300 Hz-sinusoid is added to the noise, leading to a signal whose time-frequency representation is given in the figure 4.

Analysis parameters.

The HF-HND is applied with a window length $N = 256$ samples (32ms) and an order (number of frequency bins) $K = 512$.

The HR-HND is applied with the parameters

preprocessing AR-order	filter bank	analysis (length P, order M)
order 12	no	$P = 256$ (32ms) $M = 1$

Results and interpretation. The representation of the noise part respectively extracted by the RF-HND and the HR-HND methods is displayed in the figures 5 and 6.

In both cases, a satisfying extraction of the transient noise is performed. This results from the steadiness of the sinusoidal component which matches exactly the estimated model in both cases. Nevertheless the window length cannot be shortened without increasing the variance of the HR-HND estimator or lessening the resolution capability of the RF-HND estimator. Both influences imply a notching effect on the whole extracted stochastic part, around the sinusoid frequency. A manner of this effect can be observed on the RF-noise as a "hole" in the transform around 300 Hz.

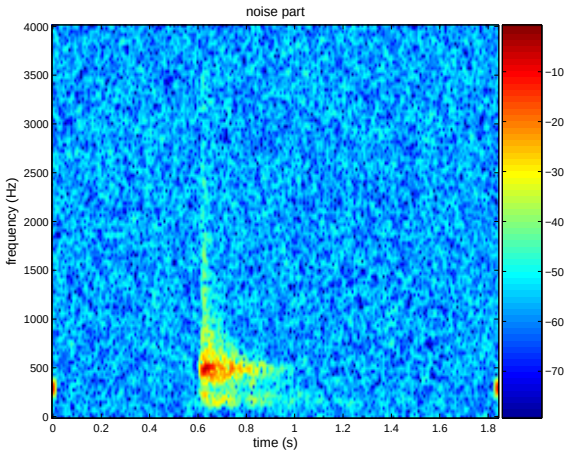

Fig. 5: Time-frequency representation of the noise part obtained by the RF-HND method.

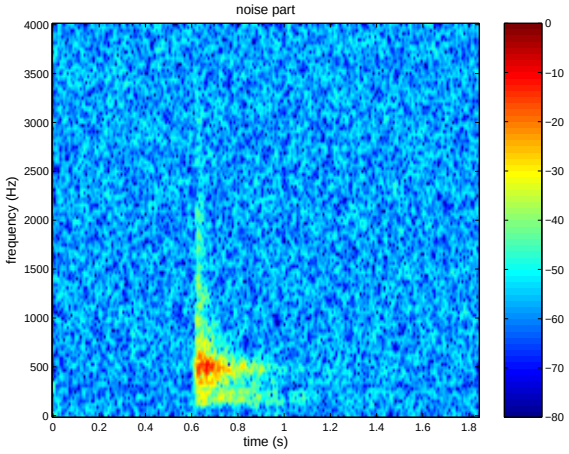

Fig. 6: Time-frequency representation of the noise part obtained by the HR-HND decomposition.

FM-modulated sine + noise. The sinusoid of this example is now modulated, leading to a 4 Hz vibrato of a semi-tone frequency deviation. Its rep-

resentation is given in figure 7.

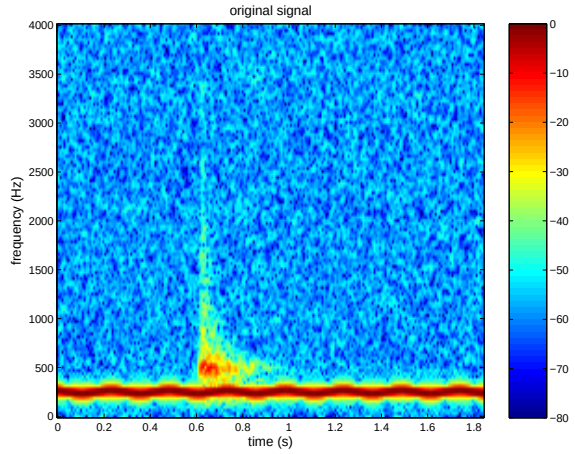

Fig. 7: Time-frequency representation of the original signal.

Analysis parameters.

The HF-HND is applied with a window length $N = 256$ samples and an order (number of frequency bins) $K = 512$.

The HR-HND is applied with the parameters

preprocessing AR-order	filter bank	analysis (length P, order M)
order 12	no	$P = 512$ (64ms) $M = 6$

Results and interpretation. The representation of the noise part respectively extracted by the RF-HND and the HR-HND methods is displayed in figures 8 and 9. This example illustrates the different tuning sensibility of both methods. For the RF-HND estimator, the window length must satisfy the trade-off between the intended resolution and the ability to track the frequency modulation. The HR-HND, in the contrary, uses a longer window and represents the modulation with the help of a higher model order. Similarly to the previous case, the spectral shape of the transient noise is roughly preserved and the RF-method causes a more pronounced notching effect.

4.3. Toward a real world application

To approach our targeted application, this last simulation is generated by mixing a real male speech utterance of the vowel 'a' (french) with the preceding transient noise.

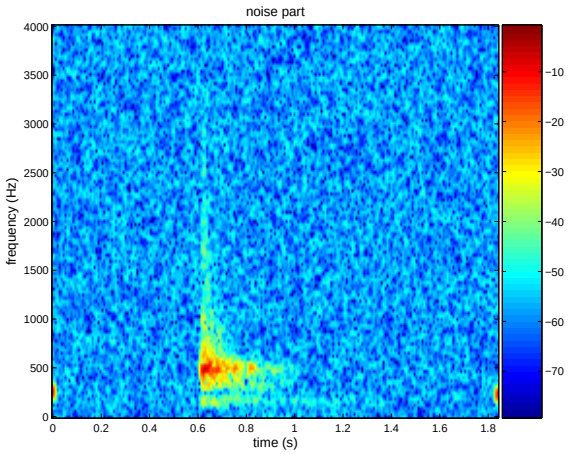

Fig. 8: Time-frequency representation of the noise part obtained by the RF-HND method.

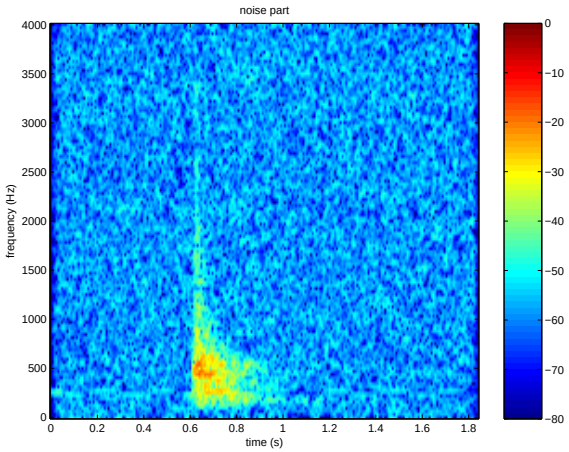

Fig. 9: Time-frequency representation of the noise part obtained by the HR-HND decomposition.

Analysis parameters.

The HF-HND is applied with a window length $N = 512$ samples and an order (number of frequency bins) $K = 1024$.

The HR-HND is applied with different parameters of the analysis for each subband.

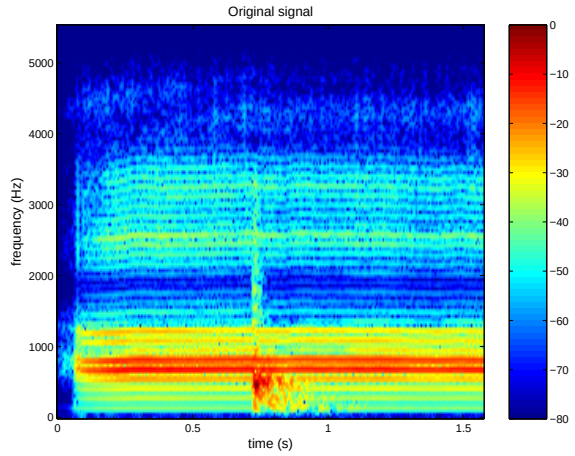

Fig. 10: Time-frequency representation of the original signal.

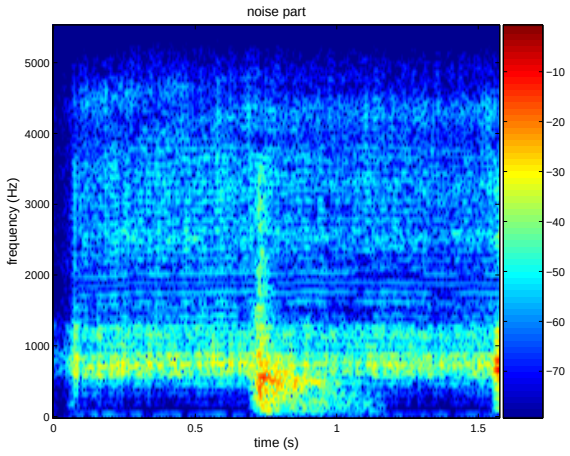

Fig. 11: Time-frequency representation of the noise part obtained by the RF-HND method.

preprocessing AR-order	filter bank	analysis (length P, order M)
order 12	4 bands AR12 whit.	$P = 200, 150, 50, 40$ $M = 40, 20, 25, 10$

Results and interpretation. The noise part extracted by the RF-HND estimator (figure 11), has a lower spectral density than that extracted by the HR-HND estimator(figure 12), especially in the upper part of the spectrum. Indeed, the subband decomposition used as preprocessing of the latter allows to process apart each subband: the window length can be adjusted differently in the lower range and in the upper range of the spectrum, leading to a kind of multiresolution processing. This might explain a more prominent notching effect for the RF-HND method while the overall formantic structure of the voice friction noise is better preserved by the HR-HND method. Conversely, the latter is more sensitive to the parameter set fine tuning,

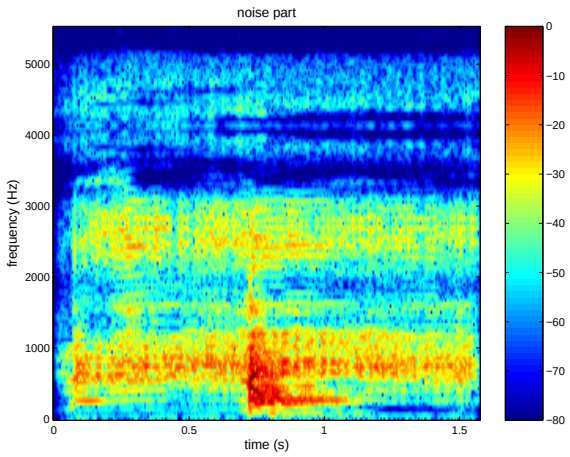

Fig. 12: Time-frequency representation of the noise part obtained by the HR-HND method.

5. CONCLUSIONS

This preliminary work on the extraction of a transient background noise surrounded by a signal with a strong harmonic content enlights the main differences and abilities of both methods: one being based on the reassigned STFT and the other being an adaptive subspace-based estimator. Both are trying to cope with the limitations related to the well-known time-frequency trade-off. Both are capable of extracting transient noise when it is reasonably strong and when the modulations of the harmonic content remain of low extent.

Future work may include the tracking of the line spectra as a preliminary of the resynthesis of the harmonic part and the test of non-uniform filter banks or multiresolution representations.

6. REFERENCES

[1] G. Peeters and X. Rodet, "SINOLA: A New Analysis/Synthesis Method using Spectrum Peak Shape Distortion, Phase and Reassigned Spectrum," in *Proc. of ICMCs*, Beijing, China, 1999.

[2] Y. Grenier and B. David, "Extraction of weak background transients from audio signals," in *Proc. of 114th Convention of the Audio Engineering Society (AES)*, Amsterdam, Netherlands, Mar. 2003.

[3] X. Serra and J. Smith, "Spectral modeling synthesis : a sound analysis/synthesis based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, 1990.

[4] L. S. Marques and L. B. Almeida, "Frequency-varying sinusoidal modeling of speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 5, pp. 763–765, May 1989.

[5] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995.

[6] C. R. Rao and L. C. Zhao, "Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals," *IEEE Trans. Signal Processing*, vol. 41, no. 3, pp. 1461–1464, Mar. 1993.

[7] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.

[8] B. G. Quinn and E. J. Hannan, *The Estimation and Tracking of Frequency*. Cambridge, UK: Cambridge University Press, 2001.

[9] M. Dvornikov, "Formulae of numerical differentiation," in *arXiv.org e-Print archive*. arXiv.org, 2003, pp. 1–14.

[10] C. Richard and R. Lengellé, "Joint recursive implementation of time–frequency representations and their modified version by the reassignment method," *Signal Processing*, vol. 60, no. 2, pp. 163–179, 1997.

[11] V. Gibiat, P. Jardin, and F. Wu, "Analyse spectrale différentielle - Application aux signaux de Myotis Mystacinus," *Acustica*, vol. 63, pp. 90–99, 1987, in French.

[12] V. Emiya, B. David, and V. Gibiat, "Two representation tools to analyse non-stationary sounds in a perceptual context," in *Proceedings of Forum Acusticum 2005*, Budapest, Hungary, Aug. 2005.

[13] D. J. Nelson, "Cross-spectral methods for processing speech," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2575–2592, 2001.

[14] J. Fourier, "Mémoire sur la propagation de la Chaleur dans les corps solides," *Nouveau Bulletin des sciences par la Société philomathique de Paris*, vol. 6, pp. 112–116, 1808, in French.

[15] G. M. Riche de Prony, "Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool différentes températures," *Journal de l'école polytechnique*, vol. 1, no. 22, pp. 24–76, 1795, in French.

[16] S. Chandran, *Advances in Direction-Of-Arrival Estimation*. Cambridge, UK: Artech House Publishers, Jan. 2006.

[17] B. David, R. Badeau, and G. Richard, "HRHATRAC Algorithm for Spectral Line Tracking of Musical Signals," in *Proc. of ICASSP'06*. Toulouse, France: IEEE, May 2006, (to be published).

[18] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysical J. Royal Astron. Soc.*, vol. 33, pp. 347–366, 1973.

[19] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[20] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT–A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1340–1342, Oct. 1986.

[21] R. Badeau, B. David, and G. Richard, "Fast Approximated Power Iteration Subspace Tracking," *IEEE Trans. Signal Processing*, vol. 53, no. 8, Aug. 2005.

[22] P. P. Vaidyanathan, *Multirate systems and filter banks*. Englewoods Cliffs, NJ, USA: Prentice Hall, 1993.