A time-frequency based method for the detection and tracking of multiple non-linearly modulated components with births and deaths

Zhongyang Li, Nadine Martin

A time-frequency based method for the detection and tracking of multiple non-linearly modulated components with births and deaths

Zhongyang Li, Nadine Martin, *Member, IEEE*

***Abstract*—The estimation of the components which contain the characteristics of a signal attracts great attention in many real world applications. In this paper, we address the problem of the tracking of multiple signal components over discrete time series. We propose an algorithm to first detect the components from a given time-frequency distribution and then to track them automatically. In the first place, the peaks corresponding to the signal components are detected using the statistical properties of the spectral estimator. Then, an original classifier is proposed to automatically track the detected peaks in order to build components over time. This classifier is based on a total divergence matrix computed from a peak-component divergence matrix that takes account of both amplitude and frequency information. The peak-component pairs are matched automatically from this divergence matrix. We propose a stochastic discrimination rule to decide upon the acceptance of the peak-component pairs. In this way, the algorithm can estimate the number, the amplitude and frequency modulation functions, and the births and the deaths of the components without any limitation on the number of components. The performance of the proposed method, a post-processing of a time-frequency distribution is validated on simulated signals under different parameter sets. The method is also applied to 4 real-world signals as a proof of its applicability.**

***Index Terms*—Time-frequency domain, multicomponent, peak detection, component tracking, amplitude and frequency modulation, nonlinear, nonstationary, births and deaths**

I. Introduction

THE detection of the multiple components of a signal and the estimation of their instantaneous amplitudes and frequencies is an important task in various domains, such as speech, fault diagnosis, civil structure analysis, and biomedical applications [1] [2] [3] [4]. In this paper, we consider a type of multicomponent signal that is particularly difficult to handle. It is characterized by the existence of multiple components with strong modulations both in amplitude and frequency. In some situations, the amplitude of certain components might diminish to or rebound from zero, which is referred to as a death or a birth of those components respectively.

Some speech processing methods allow the estimation of the amplitude and frequency modulations of multicomponent voice signals without births and deaths. SynchroSqueezing [5] decomposes the signal into a limited number of components expressed in terms of pre-defined models, namely the Intrinsic Mode Types of accuracy ϵ (ϵ-IMT). The method assumes for each ϵ-IMT a bounded chirp rate defined using ϵ to ensure the consistency with a true signal component. The signal components are reconstructed as an automatically determined number of ϵ-IMTs using the wavelet SynchroSqueezing transform that offers enhanced frequency localization compared to a non-SynchroSqueezed wavelet transform. This method has superior performance and robustness compared to the empirical mode decomposition [6], which is based on a similar component definition but lacks a rigorous theoretical interpretation.

Regarding the multicomponent nonstationary speech signals with births and deaths, spectral modeling synthesis [7] [8] is a commonly used method for speech and voice analysis. The method is based on a piecewise local spectrum analysis performed on the time-frequency distribution of the signal. The peaks, characterized by the frequency and amplitude values, are detected frame by frame using a spectrum filter, and are then tracked over each time frame using peak-component classifiers based on distance metrics, most of which measure the closeness among the frequency values. The methods in this category, such as the Harmonic plus Noise Model (HNM) [9] and the adaptive Harmonic plus Noise Model [10], are proposed to deal with the strong variations of amplitude and frequency, and even with the births and deaths [11] of voice signals. A hidden Markov model is employed in [12] to enhance the continuity of the component tracking.

Many frequency-modulation estimation methods are based on a time-frequency distribution, tool that is widely used to explore the time-varying frequency signature. The Wigner-Hough transform [13] [14] regards the Wigner-Ville distribution of the signal as an image, so that the Hough transform can be applied to detect the edge of the image as a straight line. The line parameters are suitable to characterize chirp rates. Moreover, the frequency component can either be extracted from the time-frequency distribution by using a model [15] [16] [17], or by using a peak detection method [18] [11] [8], for which the high-power frequency trajectory is selected as the instantaneous frequency. In [18], the peaks are connected by finding the nearest frequency match, whereas in [11] [7], a spectral modeling synthesis method is used to deal with more complicated frequency variations.

For long-time multicomponent signals, these methods have limitations in dealing with one or more aspects, such as the

Zhongyang Li and Nadine Martin are with Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France and CNRS, GIPSA-Lab, F-38000 Grenoble, France.

Email: zhong-yang.li@gipsa-lab.grenoble-inp.fr; nadine.martin@gipsa-lab.grenoble-inp.fr

data length, the signal types, the non-linearity types and the number of components, as well as the births and the deaths of the components. A new technique has to be developed to handle all these difficulties.

A new method is proposed in this paper, which detects and tracks the peaks of all local spectra of a given time-frequency distribution. The proposed method is also capable of estimating the amplitude modulation and the frequency modulation functions of the signal components simultaneously. As its main advantage, the method can deal with multimodulations of complicated nonlinearity types, and with the births and the deaths of the components.

Many existing nonparametric methods [19] [20] [21] are unable to estimate the amplitude modulation, many other time-frequency based methods [8] [11] [7] are able to extract the amplitude modulation functions whereas in the component tracking step, the consistency of the amplitude is not taken into account. The component tracking proposed here uses both the amplitude and the frequency values of the peaks. To deal with the extraction of the components, three questions should be primarily asked when using this approach: How can the signal be represented in the time frequency domain? How can the amplitude-frequency trajectories of a component be separated from the noise? How can the detected trajectories be connected?

The proposed algorithm is applicable on an arbitrary time-frequency distribution without cross-terms. The method proposed in this paper does not consider the reassignment of the time-frequency distribution as in Syncrosqueezing in order to avoid the disturbance introduced by the reassignment of the noise. In this paper, we illustrate the implementation for which the frequency and the amplitude of the signal components are estimated using respectively the Capongram [22] [23] and the spectrogram. These distributions are employed because the minimum variance (Capon) estimator has a low variance while the spectrogram offers a weak error in amplitude estimation.

The interpretation of the time-frequency distribution is carried out over each time frame. First, the noise spectrum, which is the spectral content corresponding to a noise that is assumed to be additive, is extracted using a data-driven method. Based on the results of the noise spectrum estimation, a peak detection method is proposed with a given false-alarm probability to distinguish the spectral content of interest from the noise. The peaks are thus detected by a Neyman-Pearson hypothesis test that requires the choice of a false-alarm probability and the estimation of the noise spectrum.

Using a component-tracking method, the detected peaks are connected in sequence to form the amplitude and frequency modulations of the components. The component-tracking method, as proposed in [18], where only a few neighboring time-frequency samples are used, is hardly applicable when births and deaths occur, or when the frequency variation is strong. To deal with the strong nonstationarity of the signal, we propose to use a peak-component classification in order to track the amplitude and frequency modulations inspired by the K-means classifier [24] and by the Spectral modeling synthesis methods [11] [7]. The discrimination criterion of the classification yields three possible results: acceptance, rejection, and creation of new components. A weighted Euclidean distance of amplitude and frequency is used as the distance metric between a peak and a component. In this manner, the tracking strategy is adaptive and robust.

The content of this paper is organized as follows. Section II gives the general definition of a signal of interest and presents the proposed algorithm in detail. The performance of the proposed method is analyzed with simulated data in Section III, and illustrated on real-world signals in Section IV. Finally, conclusions are drawn in section V.

II. The proposed algorithm

The model of the complex observation $y[n]$ is given as,

$$\begin{aligned} y[n] &= s[n] + e[n], \qquad 0 \le n \le N-1 \\ s[n] &= \sum_{k=1}^{K} I_k[n] A_k[n] e^{j\Phi_k[n]}, \end{aligned} \tag{1}$$

where n is the discrete time index, $s[n]$ is a deterministic noise-free signal, $e[n]$ is a white gaussian noise of zero mean and unknown variance σ^2, and K is the number of components of the signal. $I_k[n]$ is an indicator function that indicates the existence of component k,

$$I_k[n] = \begin{cases} 1 & \text{if component } k \text{ exists at } n, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

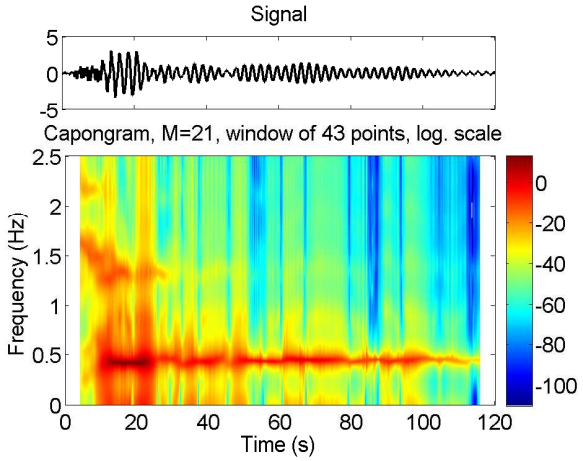
$A_k[n]$ and $\Phi_k[n]$ are the instantaneous amplitude and phase of the k-th component. We define the instantaneous frequency as the discrete derivation of $\Phi_k[n]$,

$$F_k[n] = \frac{1}{2\pi}(\Phi_k[n+1] - \Phi_k[n]) I_k[n], \tag{3}$$

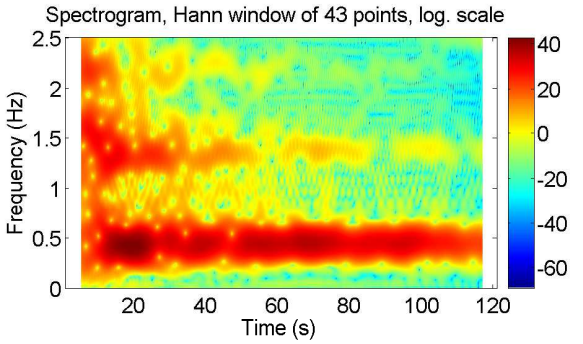
When component k exists, we assume $A_k[n]$ to be positive and $\Phi_k[n]$ to be continuous [25] [26]. The frequency value is continuous, positive and bounded $0 < F_k[n] < F_s/2$, where F_s is the sampling frequency. In this paper, the frequency estimation is based on a time-frequency distribution of $N_{\mathcal{Y}}+1$ frequency bins, such that $F_k[n] \in \{0, F_s/2N_{\mathcal{Y}}, \ldots, f \times F_s/2N_{\mathcal{Y}}, \ldots, F_s/2\}$.

However, when component k dies, $\Phi_k[n]$ is not defined, and therefore the derivation $(\Phi_k[n+1] - \Phi_k[n])$ is unknown. The above-mentioned Bedrosian identity is not applicable. Using the indicator function $I_k[n]$, the instantaneous frequency during the component death is set to zero, to avoid ambiguity with its physical interpretation.

In this section, we describe the main steps of the proposed algorithm based on a time-frequency distribution $\mathcal{TFD}_y[n, f]$ of the signal $y[n]$. The calculation of the latter depends on the user and should be tailored according to the application. This choice is beyond the scope of this paper. In the following sections, the results of the proposed method are demonstrated using two time-frequency distributions. The first one is the spectrogram which provides a piece-wise Fourier-based power spectral density estimate. The second one is the Capongram, which is a piece-wise power spectrum obtained using a non-Fourier estimation method, called the minimum variance method [27] [28]. The details of these time-frequency distributions are presented in the appendix. An illustration is given in Fig.1.



(a) Capongram of a seismic signal



(b) Spectrogram of a seismic signal

Fig. 1. Two time-frequency distributions of a seismic signal of 120 s ($n = 1, \ldots, 601$), resampled at 5 Hz, with $N_{\mathcal{Y}} = 505$ frequency bins. (a) Capongram with a window of 43 points (8.6 s); (b) Spectrogram with a Hann window of 43 points (8.6 s).

The proposed method consist of a peak detection step and a component tracking step. The peaks of the local spectrum of time instant n, noted as $\mathcal{Y}[n] = \mathcal{TFD}_y[n, f]$ are detected in two steps. The first step is to estimate the additive noise spectrum. The estimation is done by a multipass filtering technique, as described in Section II-B. The second step is to detect the peaks of the local spectra, using a data-driven peak detection method, as described in Section II-C. Both of these steps are based on a Neyman-Pearson hypothesis test, as described in Section II-A. The detected peaks are finally connected by the component tracking method proposed in Section II-D.

A. Hypothesis test for peak detection and removal

Since the additive noise $e[n]$ is of zero mean, the cross-section of the time-frequency distribution at time instant n can be expressed as

$$\mathcal{Y}^{(n)}[f] = \mathcal{S}^{(n)}[f] + \mathcal{E}^{(n)}[f], \tag{4}$$

where $\mathcal{Y}^{(n)}[f] = \mathcal{TFD}_y[n, f]$. $\mathcal{S}^{(n)}[f]$ and $\mathcal{E}^{(n)}[f]$ are the cross-sections of respectively the deterministic signal $s[n]$ and the additive noise $e[n]$. We define here two types of spectral content as a double-hypothesis test,

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathcal{Y}^{(n)}[f] = \mathcal{E}^{(n)}[f], \\ \mathcal{H}_1 : &\quad \mathcal{Y}^{(n)}[f] = \mathcal{S}^{(n)}[f] + \mathcal{E}^{(n)}[f]. \end{aligned} \tag{5}$$

When $\mathcal{H}_1$ is unknown, as is the case in this paper, we can apply a Neyman-Pearson lemma [1], with

$$\mathcal{PFA} = Pr(T(f) > \lambda | \mathcal{H}_0), \tag{6}$$

where λ is the test threshold, and $\mathcal{PFA}$ is the false-alarm probability. The test result is given by

$$T(f) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \tag{7}$$

In this test, the threshold λ is calculated from a chosen $\mathcal{PFA}$ and using the statistical properties of the spectrum estimator. For Capon and Fourier estimators,

$$T(f) = \frac{\mathcal{Y}^{(n)}[f]}{\mathcal{E}^{(n)}[f]}. \tag{8}$$

Under the hypothesis $\mathcal{H}_0$

$$2T(f) \sim \chi^2_2. \tag{9}$$

$T[f]$ can be regarded as proportional to a random variable which has a χ^2_2 distribution [28] if the noise $e[n]$ is gaussian. Combining (6) and (8), the detection threshold λ can be determined by

$$\mathcal{PFA} = \int_{2\lambda}^{+\infty} p_{T(f)|\mathcal{H}_0}(x)dx = \int_{2\lambda}^{+\infty} p_{\chi^2_2}(x)dx. \tag{10}$$

Then the test hypothesis (7) becomes

$$\mathcal{Y}^{(n)}[f] \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda \times \mathcal{E}^{(n)}[f]. \tag{11}$$

From (11), the test depends on the *a priori* choice of the false-alarm probability $\mathcal{PFA}$ and the estimation of the unknown noise spectrum $\mathcal{E}^{(n)}[f]$.

In the proposed algorithm, the test is used twice: in the noise spectrum estimation and in the peak detection. For the noise spectrum estimation described in Section II-B, it is necessary to remove peaks using this hypothesis test with a false-alarm probability referred to as $\mathcal{PFA}_e$. For the peak detection in Section II-C, the test is applied using another value, referred to as $\mathcal{PFA}_d$.

B. Estimation of the noise spectrum

Focusing on a given time frame n, we propose to approximate the unknown noise spectrum $\mathcal{E}^{(n)}[f]$ through an iterative process; namely, the multipass filtering that was developed in [29] [30]. The entire procedure of multipass filtering can be summarized by Fig.2.

The first pass is a L_f-point median filter,

$$\hat{\mathcal{E}}^{(n)}_1[f] = \mathbf{Median}\{\mathcal{Y}^{(n)}[f - \frac{L_f - 1}{2}], \ldots, \mathcal{Y}^{(n)}[f + \frac{L_f - 1}{2}]\}, \tag{12}$$

where L_f is the length of a frequency window sliding over the frequency samples. $\hat{\mathcal{E}}_1^{(n)}[f]$ is the estimation result of the first pass.

The estimated noise spectrum is refined in the following passes. Each pass p has two steps: in the first step, the peaks of the spectrum $\mathcal{Y}^{(n)}[f]$ are removed by applying the hypothesis test described in the previous section, with a false-alarm probability $\mathcal{PFA}_e$ and the previously estimated noise spectrum $\hat{\mathcal{E}}_{p-1}^{(n)}[f]$. All peaks verifying $\mathcal{H}_1$ are removed from the local spectrum to yield its peak-free part $E_p^{(n)}[f]$. The second step smooths the remaining part $E_p^{(n)}[f]$ by an average filter with a sliding window of L_f points. The estimated noise spectrum is upgraded by the filter output. An example of the first and last passes is shown in Fig.3.

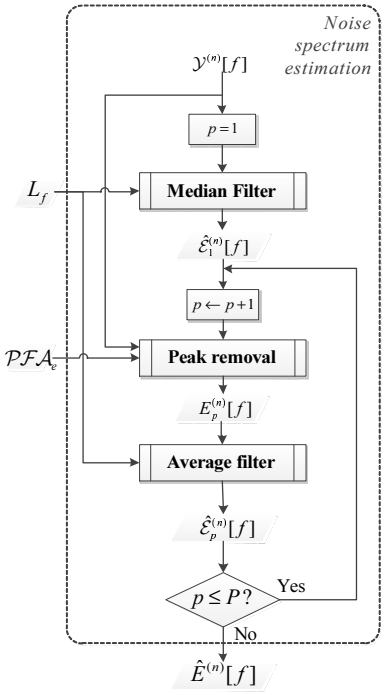
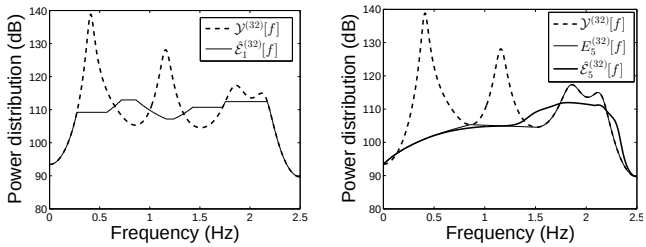


Fig. 2. Decomposition of the block "noise spectrum estimation" in Fig.9, done by multipass filtering with P passes. Based on the local spectrum $\mathcal{Y}^{(n)}[f]$, the noise spectrum is finally estimated as $\hat{\mathcal{E}}^{(n)}[f]$ using a sliding window of L_f frequency samples and the probability of false alarm $\mathcal{PFA}_e$.



(a) first pass (b) second pass

Fig. 3. The noise spectra of the local spectrum extracted from the Capongram of Fig.1(a) at $n = 32$, obtained by the first pass ($p = 1$) and the last pass ($p = P, P = 5$) from the original local spectrum $\mathcal{Y}^{(32)}[f]$, $f = 1, \ldots, N_{\mathcal{Y}}$, $N_{\mathcal{Y}} = 505$. (a) The first pass: $\hat{\mathcal{E}}_1^{(32)}[f]$ calculated by the median filtering of 169 points; (b) The last pass: $\hat{\mathcal{E}}_5^{(32)}[f]$ calculated by the average filtering of 169 points of $E_5^{(32)}[f]$ obtained with $\mathcal{PFA}_e = 0.1$.

The final noise spectrum $\hat{\mathcal{E}}^{(n)}[f]$ estimation is the output after P passes,

$$\mathcal{E}^{(n)}[f] \approx \hat{\mathcal{E}}^{(n)}[f] = \hat{\mathcal{E}}_P^{(n)}[f]. \tag{13}$$

The total number of passes P corresponds to the pass from which the estimated noise spectrum remains identical.

C. Peak detection

The estimated noise spectrum is regarded as the spectral content verifying $\mathcal{H}_0$ in (5). Based on the estimation result, another Neyman-Pearson test is applied to identify the peaks that verify $\mathcal{H}_1$. For each frequency index f, a hypothesis test is carried out using a statistical test of the random variable $T[f]$

$$T(f) = \frac{\mathcal{Y}^{(n)}[f]}{\hat{\mathcal{E}}^{(n)}[f]}, \tag{14}$$

under the hypothesis $\mathcal{H}_0$

$$2 \times T(f) \sim \chi_2^2. \tag{15}$$

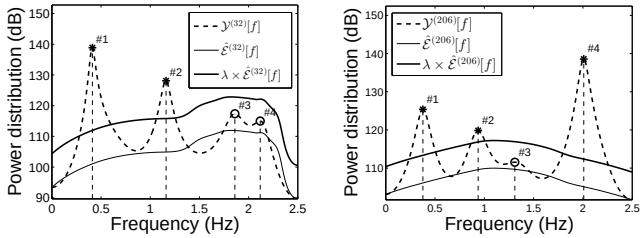
The test threshold λ is defined by a false-alarm probability $\mathcal{PFA}_d$ as

$$\mathcal{PFA}_d = \int_{\lambda}^{+\infty} p_{T(f)|\mathcal{H}_0}(x)dx = \int_{2\lambda}^{+\infty} p_{\chi_2^2}(x)dx. \tag{16}$$

Then the hypothesis of peak detection (7) becomes

$$\mathcal{Y}^{(n)}[f] \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda \times \hat{\mathcal{E}}^{(n)}[f]. \tag{17}$$

An example of the peak detection is shown in Fig.4.



(a) $n = 32$ (b) $n = 206$

Fig. 4. Peak detection on two local spectra of the signal of Fig.1 at $n = 32$ and $n = 206$, with $\mathcal{PFA}_e = \mathcal{PFA}_d = 0.1$. (*)Detected peaks over the threshold $\lambda \times \hat{\mathcal{E}}^{(n)}[f]$ calculated with $\mathcal{PFA}_e = 0.1$ and $\hat{\mathcal{E}}^{(n)}[f]$. (○) Local maxima treated as noise.

The peaks are defined as the vertex of the lobes above the detection threshold $\lambda \times \hat{\mathcal{E}}^{(n)}[f]$. Each peak is represented by its frequency $f_q^{(n)}$ and its amplitude $a_q^{(n)}$. Due to the nonstationarity of the signal, the number $Q^{(n)}$ of peaks varies with time n (for example in Fig.4(a) the number of detected peaks $Q^{(32)} = 2$ at time frame $n = 32$, whereas in Fig.4(b), $Q^{(206)} = 3$ at time frame $n = 206$).

D. Component tracking

We hereby address the issue of component tracking from the detected peaks, as summarized in Fig.5.

Suppose that all the components remain inactive until the time frame n_{start} when the first $Q^{(n_{start})} = Q_{(n)}|_{n=n_{start}}$ peaks are detected. The peaks are directly stored as the beginning of $Q^{(n_{start})}$ components,

$$\begin{aligned} &\hat{F}_k[n_{start}] = f_k^{(n_{start})}\tfrac{F_s}{2N_\mathcal{Y}}, \hat{A}_k[n_{start}] = a_k^{(n_{start})}, \\ \text{for} \quad &k \le \hat{K}_{n_{start}}, \\ \text{where} \quad &\hat{K}^{(n_{start})} = Q^{(n_{start})}, \end{aligned} \tag{18}$$

where $\hat{F}_k[n]$ and $\hat{A}_k[n]$ denote the estimation of the frequency modulation $F_k[n]$ and of the amplitude modulation $A_k[n]$ of the k-th component.

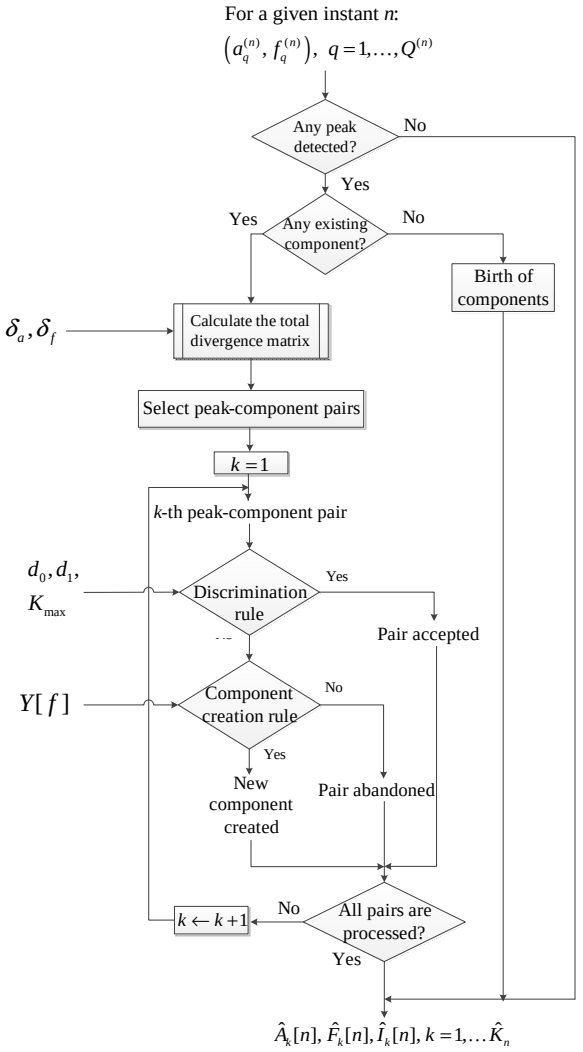


Fig. 5. Flow chart of the component tracking. The $Q^{(n)}$ detected peaks are classified into $\hat{K}_n$ components to formulate the amplitude and frequency modulations $\hat{A}_k[n]$ and $\hat{F}_k[n]$ and the indicator functions $\hat{I}_k[n]$, using the parameters δ_a, δ_f, d_0, and d_1.

For the following $n > n_{start}$ time frames where the tracked components are non-empty, the component tracking is a sequential and iterative procedure to associate $\hat{K}^{(n-1)}$ previously built components with $Q^{(n)}$ detected peaks using a peak-component classifier. The principle of tracking is to choose the peak-component pairs with closest amplitude and frequency values. Therefore, the classification at each time frame n starts with calculating the distances of all the available peak-component pairs and selecting the pairs with the smallest distances. This is done by the total divergence matrix we propose and describe in Section II-D1. Then the selected pairs are accepted or rejected according to a discrimination rule defined in Section II-D2, or otherwise regarded as the beginning of new components using a criterion described in Section II-D3.

1) Total divergence matrix: The peak-component classifier is based on two essential elements: the peaks are objects to be classified, while the components are classes that store a similar group of peaks. Differently from using a canonical classifier, the peaks are stored sequentially, and at each time frame n one component can associate with only one "optimal" peak that is closest in both amplitude and frequency. Thus the pairs have to be selected by means of what we refer to as the total divergence matrix, which gathers all the possible choices.

The distance for each peak-component pair has first to be computed. Since the quality of selection relies totally on the quality of the distance calculation, we need to take both amplitude and frequency values into account. The amplitude and frequency of component k are characterized by $\hat{A}_k[n'_k]$ and $\hat{F}_k[n'_k]$ at time frame n'_k, which corresponds to the most recent non-zero value of $\hat{I}_k[i] = 0$,

$$\hat{I}_k[n'_k] > 0 \quad \text{and } \hat{I}_k[i] = 0, \text{ for } n'_k \le i < n, \tag{19}$$

while the amplitude and frequency of a detected peak q is defined as $a_q^{(n)}$ and $f_q^{(n)}\frac{F_s}{2N_\mathcal{Y}}$. The amplitude distance $\Delta a_{k,q}$ and frequency distance $\Delta f_{k,q}$ between component k and peak q at time frame n is defined as

$$\begin{aligned} \Delta a_{k,q} &= |\hat{A}_k[n'_k] - a_q^{(n)}|, \\ \Delta f_{k,q} &= |\hat{F}_k[n'_k] - f_q^{(n)}\tfrac{F_s}{2N_\mathcal{Y}}|. \end{aligned} \tag{20}$$

Such distance values are calculated for all the $\hat{K}^{(n-1)}$ existing components and all the $Q^{(n)}$ detected peaks. We define an amplitude divergence matrix $\mathbf{\Delta A}$ and a frequency divergence matrix $\mathbf{\Delta F}$, both of dimension $\hat{K}^{(n-1)} \times Q^{(n)}$, such as

$$\begin{aligned} \mathbf{\Delta A} &= [\Delta a_{k,q}]_{\hat{K}^{(n-1)} \times Q^{(n)}}, \quad \Delta a_{k,q} = |\hat{A}_k[n'_k] - a_q^{(n)}|, \\ \mathbf{\Delta F} &= [\Delta f_{k,q}]_{\hat{K}^{(n-1)} \times Q^{(n)}}, \quad \Delta f_{k,q} = |\hat{F}_k[n'_k] - f_q^{(n)}\tfrac{F_s}{2N_\mathcal{Y}}|. \end{aligned} \tag{21}$$

An example of an amplitude and a frequency divergence matrix is illustrated in Table I.

TABLE I
THE AMPLITUDE AND FREQUENCY DIVERGENCE MATRIX OF PEAKS $q = 1, 2$ OF FIG.4 AND TWO COMPONENTS $k = 1, 2$ ALREADY TRACKED AT $n = 32$ ON THE CAPONGRAM OF FIG.1(A) USING $\mathcal{PFA}_d = \mathcal{PFA}_e = 0.1$. THE TRACKING CANDIDATES ARE IN BOLD.

Elements of ΔF and ΔA			peak q	
			1	2
component k	1	$\Delta f_{k,q}$	**0**	0.75
		$\Delta a_{k,q}$	**5.44**	25.89
	2	$\Delta f_{k,q}$	0.75	**0.005**
		$\Delta a_{k,q}$	20.34	**0.04**

To facilitate the selection of peak-component pairs, the two matrices are merged into a "total divergence matrix" according to the following equation,

$$\begin{aligned} &\mathbf{\Delta D} = [\Delta d_{k,q}]_{\hat{K}_{n-1} \times Q^{(n)}}, \\ \text{with} \quad &\Delta d_{k,q} = \sqrt{\Delta \overline{a}_{k,q}^2 + \Delta \overline{f}_{k,q}^2}, \\ \text{and} \quad &\Delta \overline{a}_{k,q} = \delta_a \times \tfrac{\Delta a_{k,q}}{\Delta a_{max}}, \Delta \overline{f}_{k,q} = \delta_f \times \tfrac{\Delta f_{k,q}}{\Delta f_{max}}. \end{aligned} \tag{22}$$

where Δa_{max} and Δf_{max} are two normalization factors, *i.e.* the maximum values of the elements of $\mathbf{\Delta A}$ and $\mathbf{\Delta F}$. δ_a and δ_f are the amplitude and the frequency weighting factors that satisfy $\delta_a^2 + \delta_f^2 = 1$. A greater frequency weighting factor means that the frequency divergence is magnified while the amplitude divergence is suppressed, and *vice versa*. For many real-world signals, the choice of these two factors is quite flexible. The total divergence matrix $\mathbf{\Delta D}$ calculated by synthesizing Table I using (22) is shown in Table II using $\delta_f = 2\delta_a$ and $\delta_f = \delta_a$.

TABLE II
THE TOTAL DIVERGENCE MATRIX SYNTHESIZED FROM TABLE I ACCORDING TO (22) USING TWO CONFIGURATIONS: $\delta_f = \delta_a$ AND $\delta_f = 2\delta_a$. THE TRACKING CANDIDATES ARE IN BOLD.

Elements of the total divergence matrix			peak q	
			1	2
component k	1	$\delta_f = 2\delta_a$	***0.21***	1.41
		$\delta_f = \delta_a$	***0.12***	1.41
	2	$\delta_f = 2\delta_a$	*1.26*	**0.006**
		$\delta_f = \delta_a$	*1.35*	**0.008**

The selection of the peak-component pairs matches each component k with the best peak q_k, by locating the smallest total divergence value of the row k in the total divergence matrix $\mathbf{\Delta D}$,

$$q_k = \underset{1 \le q \le Q^{(n)}}{\arg\min} \{\Delta d_{k,q}\}. \tag{23}$$

Once the index q_k is found for the component k, all the elements on row k and column q_k are removed from the total divergence matrix $\mathbf{\Delta D}$ to form a submatrix. The next pair is selected by finding the minimum value in the submatrix. In this manner, every existing component is matched with a unique peak; the selection continues until the submatrix reduces to a vector.

2) Discrimination rule: A peak-component pair thus selected might be wrong if the total divergence value is high, because false peaks might exist because of noise and interference terms of the time-frequency distribution. A discrimination rule based on the total divergence value is necessary to judge whether a pair can be accepted. Due to the random disturbance of the noise, we assume that a pair that is correctly associated should have a low total divergence value that satisfies a stochastic discrimination rule; otherwise, it is regarded as a wrong association. This discrimination rule is defined as:

(i) Generate a random variable u by a uniform distribution $\mathcal{U}[0, 1]$;
(ii) Accept the association if $u < \eta(\Delta d_{k,q_k})$; otherwise, the association is discarded.

This can be also expressed as

$$(\hat{A}_k[n], \hat{F}_k[n]) = \begin{cases} (a_{q_k}^{(n)}, f_{q_k}^{(n)} \times \frac{F_s}{2N_{\mathcal{Y}}}) & \text{if } u < \eta(\Delta d_{k,q_k}), \\ (0, 0) & \text{otherwise,} \end{cases} \tag{24}$$

where $\Delta d_{k,q_k}$ is the total divergence value of the selected pair (k, q_k). To determine whether the pair is to be accepted or discarded, a threshold function $\eta(\Delta d)$ of total divergence Δd is applied on $\Delta d_{k,q_k}$ as

$$\eta(\Delta d_{k,q_k}) = \begin{cases} 1 & \text{if } 0 \le \Delta d_{k,q_k} < d_0, \\ exp(-\frac{1}{2}(\frac{\Delta d_{k,q_k} - d_0}{\sigma(1-d_0)})^2) & \text{if } \Delta d_{k,q_k} \ge d_0, \end{cases} \tag{25}$$

where $\sigma = 2\sqrt{2}\frac{1-d_0}{d_1-d_0}$ is a constant designating $\eta(d_1) = 10^{-4}$.

The function η indicates the tolerance of the divergence using two parameters d_0 and d_1. Parameter d_0 is a lower threshold, in the sense that a pair with a total divergence value lower than d_0 is definitely accepted. $(d_1 - d_0)$ influences the rate of acceptance of a pair with total divergence value higher than d_0 using the random test based on u. A high value of $(d_1 - d_0)$ results in a wider region of acceptance.

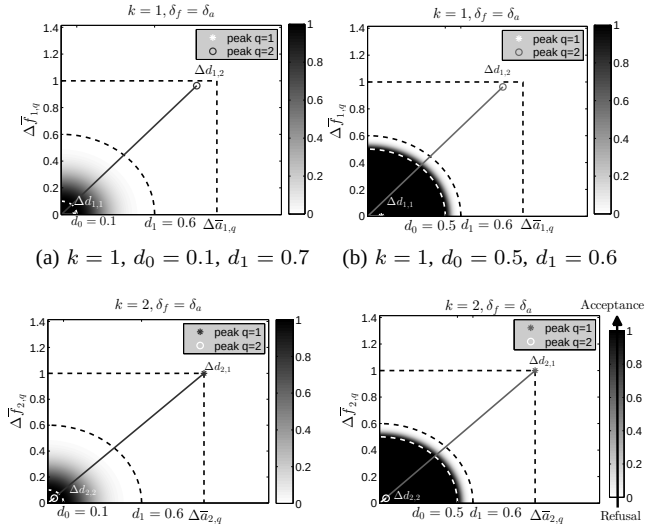
Parameter d_1 controls the tolerance of the total divergence. Successful component tracking also depends on δ_a and δ_f in (22), which control the relative importance of the amplitude and frequency divergence. The impact of δ_a, δ_f, d_0 and d_1 is shown in Fig.6 and Fig.7.

η therefore defines a soft threshold for the discrimination rule. Instead of either accepting or rejecting the association by applying a hard threshold on the total divergence value, η defines a smooth transition where the discrimination can be decided by a probability, specified by d_0 and d_1. When $d_1 = 0$, the discrimination rule reduces to a classical hard thresholding.

In Fig.6, the total divergence value $\Delta d_{k,q}$ is illustrated between two components $k = 1, 2$ and two peaks $q = 1, 2$. Each total divergence value is represented by a two-dimension vector. The darker the region a total divergence value locates in, the higher the level of acceptance of the corresponding peak-component pair. A higher d_0 value defines a wider dark region (see Fig.6(a) and Fig.6(b)), whereas a larger $d_1 - d_0$ induces a smoother transition outside. The value d_0 is important for the pairs $k = 1$, $q = 1$ and $k = 2$, $q = 2$, which have small total divergence values (see Fig.6(a) and Fig.6(c)). Nonetheless, the pair $k = 1$, $q = 2$, with a high total divergence value that corresponds to a wrong association, is

almost impossible to be accepted since $\eta(\Delta d_{1,2})$ is very close to 0. In Fig.6, it is evident that in case of $k = 1$ as shown in (a) and (b), the peak of the best match $q_1 = 1$. In another case where $k = 2$ as shown in (c) and (d), the peak of the best match $q_2 = 2$. The gap defined by $d_1 - d_0$ affects the acceptance of the pairs outside. Indeed, the values of d_0 and d_1 are not very sensitive unless they become close to 0 or 1.

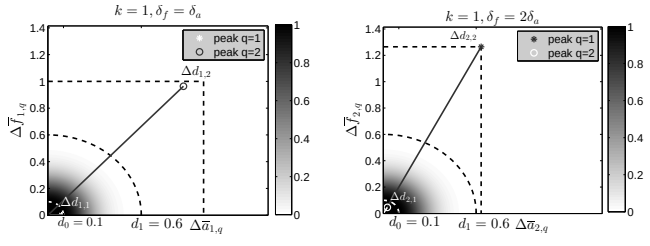
From Fig.7(a) to Fig.7(b), the total divergence values are influenced by the increase in δ_f compared to δ_a. The contribution of the amplitude divergence is suppressed with respect to the frequency divergence.



(a) $k = 1$, $d_0 = 0.1$, $d_1 = 0.7$ (b) $k = 1$, $d_0 = 0.5$, $d_1 = 0.6$

(c) $k = 2$, $d_0 = 0.1$, $d_1 = 0.6$ (d) $k = 2$, $d_0 = 0.5$, $d_1 = 0.6$

Fig. 6. Impact of d_0 and d_1 on the tolerance of the total divergence, illustrated with data of Table I under $\delta_f = \delta_a$. d_0 is equal to the radius of the inner circle, d_1 is equal to the diameter of the outer circle. The function $\eta(\Delta d_{k,q})$, which is represented by the gray level of the image, changes accordingly. δ_a is equal to the width of the outer rectangle, δ_f is equal to its height. The length of the vectors indicates the corresponding total divergence value $\Delta_{k,q}$ in Table II, where q is distinguished by different gray scales.



(a) $k = 1$, $\delta_f = \delta_a$ (b) $k = 1$, $\delta_f = 2 \times \delta_a$

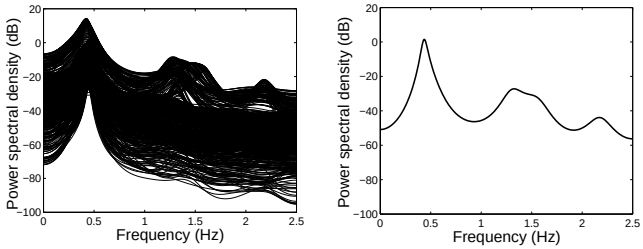
Fig. 7. Impact of δ_a and δ_f illustrated with data of Table I. δ_a is equal to the width of the outer rectangle, and δ_f is equal to its height. The length of the vectors indicates the corresponding total divergence value $\Delta d_{k,q}$ in Table II, where q is distinguished by different gray scales.

3) Creation of new components: If a pair is rejected by the above discrimination rule, the peak q might be the beginning of a new component. We presume that the peak is likely to be a component if it locates in the energetic part of the global spectrum, as illustrated in Fig.8. An energetic part of the local spectra corresponds to an energetic part of the global spectrum. This means that the creation of a new component can be determined by using the global spectrum.

The creation of a component yields a component activated at instant n, expressed as,

$$\begin{aligned} &\hat{K}^{(n)} = K^{(n-1)} + 1, \hat{F}_{\hat{K}^{(n)}}[i]|_{i<n} = 0, \\ &\hat{A}_{\hat{K}^{(n)}}[i]|_{i<n} = 0, \hat{F}_{\hat{K}^{(n)}}[n] = f_q \times \tfrac{F_s}{2N_{\mathcal{Y}}}, \hat{A}_{\hat{K}^{(n)}}[n] = a_q, \\ &\text{if} \qquad \mathcal{Y}[f_q] > \lambda \times \hat{\mathcal{E}}[f], \end{aligned} \tag{26}$$

where $\hat{\mathcal{E}}[f]$ is the global noise spectrum estimated using the multipass filtering proposed in Section II-B with false-alarm probability $\mathcal{PFA}_e$ and number of passes P; λ is a detection threshold that is calculated using the method proposed in Section II-C with false-alarm probability $\mathcal{PFA}_d$.



(a) (b)

Fig. 8. (a) Superimposed local spectra of the Capongram $\mathcal{Y}^{(n)}[f]$ of Fig.1(a) calculated using a Capon estimator at order 21, with a 43-point window and 505 frequency bins. (b) $\hat{\mathcal{E}}[f]$: Spectrum of the entire signal of Fig.1, calculated using a Capon estimator at order 21 and 505 frequency bins.

E. Summary

The above-described steps are arranged in the proposed algorithm as illustrated in Figure 9.

One of the advantages of the proposed algorithm is that the peak detection strategy is adjusted to the data and the statistical properties of the spectrum estimator. Therefore, heavy manual configurations are avoided. With the noise spectrum $\hat{\mathcal{E}}^{(n)}[f]$ estimated from the local spectrum itself using a multipass filtering method, the calculation of the detection threshold depends on the false-alarm probability chosen by the user. The spectral contents of interest hereby detected are referred to as "peaks", which are essential to the component tracking. Each peak refers to a pair of amplitude and frequency values. This peak-detection strategy is particularly suitable for highly nonstationary signals for which a manual choice of detection threshold is almost impossible over each time frame.

As another contribution, we propose a component-tracking method to upgrade the amplitude and frequency modulations using the amplitude and frequency values of the detected peaks. The tracking method automatically determines which component a peak belongs to and how the components should be tracked. Based on the hypothesis that if two peaks can be connected as one component, their frequency and amplitude distances should simultaneously be very low, these differences are evaluated by a proposed distance metric; namely, the total divergence. The judgement on whether two peaks can be connected is carried out by a peak-component classifier. The component is considered dead if no further peak can be found.

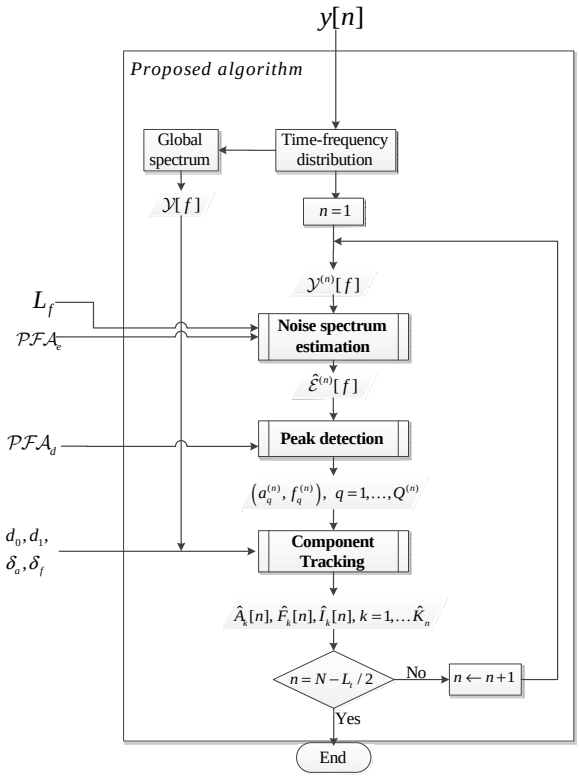


Fig. 9. Flow chart of the proposed algorithm. Given the input signal $y[n]$, the algorithm provides the estimated number of components $\hat{K}_{N-L_t/2}$, the amplitude and the frequency modulations $\hat{A}_k[n]$, $\hat{F}_k[n]$, and the indicator functions $\hat{I}_k[n]$, using the parameters L_t, $\mathcal{PFA}_d$, P, $\mathcal{PFA}_e$, δ_a, δ_f, d_0, and d_1.

In this way, the births and the deaths of components are also managed.

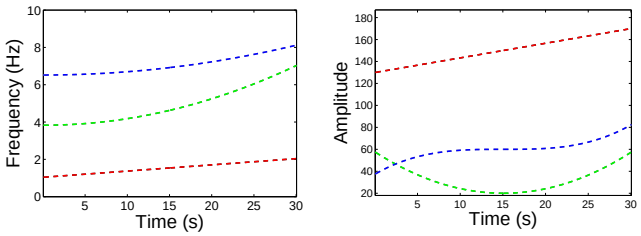
III. Performance Analysis on Simulated Signals

In this section, the performance of the algorithm will be assessed under various aspects. The performance of the frequency and the amplitude estimation is evaluated over multicomponent signals without births or deaths under different signal-to-noise ratio (SNR) levels. The recognition of the births and the deaths of the components is then studied in three cases: births or deaths that occur on all components simultaneously, at different time frames, or at very short time intervals. An advantage of this nonparametric approach is that the number of components to handle is not limited by the dimension of the parameter space.

A. On multicomponent signals without births or deaths

In this section, we analyze the performance of the proposed algorithm on multicomponent simulated signals under different parameter sets. Among the algorithm parameters indicated in Fig.9, $\mathcal{PFA}_e$, $\mathcal{PFA}_d$, δ_a, δ_f, d_0, and d_1 have to be chosen manually by the user. The sensitivity of these parameters is studied in order to assess the robustness of the algorithm.

In real-world situations, the instantaneous amplitudes of the components can be relatively disparate; therefore, the components are generated with diversified amplitude levels in order to be close to reality. Consider first the case without births or deaths, where zero values do not exist in the amplitude and the frequency modulations. The simulated signal $y[n]$ is defined as (1), with $N = 600$ samples, sampling frequency $F_s = 20Hz$, and $K = 3$ components. The amplitude and the frequency modulations of $s[n]$ are generated as shown in Fig.10.



(a) Frequency modulations (b) Amplitude modulations

Fig. 10. Generated frequency and amplitude modulations of a three-component simulated signal, where the components are distinguished by different colors ($k = 1$,$k = 2$,$k = 3$).

The global SNR is set to $15dB$, defined as the ratio of the average power of the signal $\sum_{k=1}^{K} |s[n]|^2$ to the noise variance σ_e^2. For a highly nonstationary signal, the global SNR can hardly characterize the local noise level against individual components. For this signal, the power of the strongest component is over 50 times higher than the weakest one at $n = 298$. Fig.11 illustrates this effect for each component by showing its local SNR, which is the ratio of its instantaneous power $A_k^2[n]$ to the noise variance σ_e^2.

Even for a moderate global SNR level, the components $k = 1$ and $k = 3$ are rather distinct, as shown by the Capongram in Fig.12, while the component $k = 2$ is almost as weak as the noise for a local SNR equal to -5 dB.

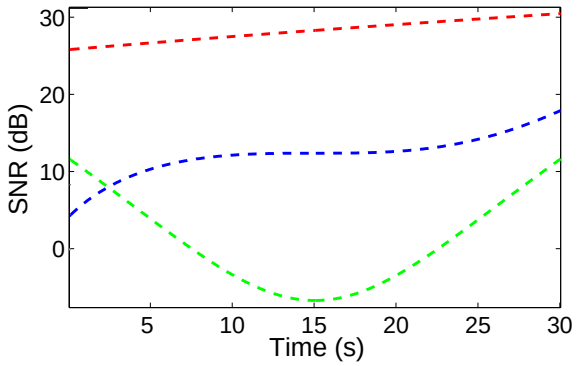


Fig. 11. Time-varying local SNRs of the simulated signal with the three components defined in Fig.10, at a global SNR of 15 dB. The components are distinguished by different colors ($k = 1$,$k = 2$,$k = 3$).

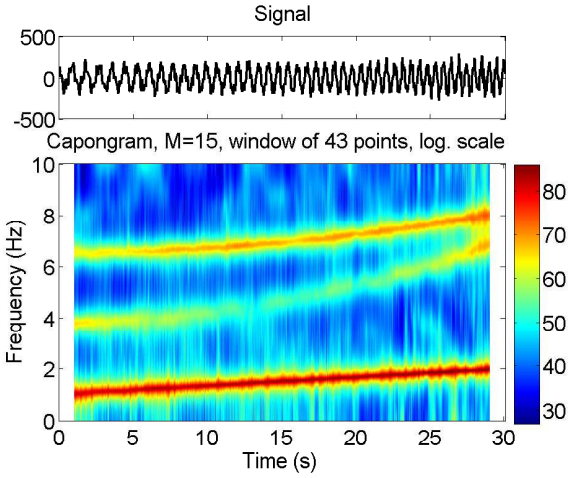
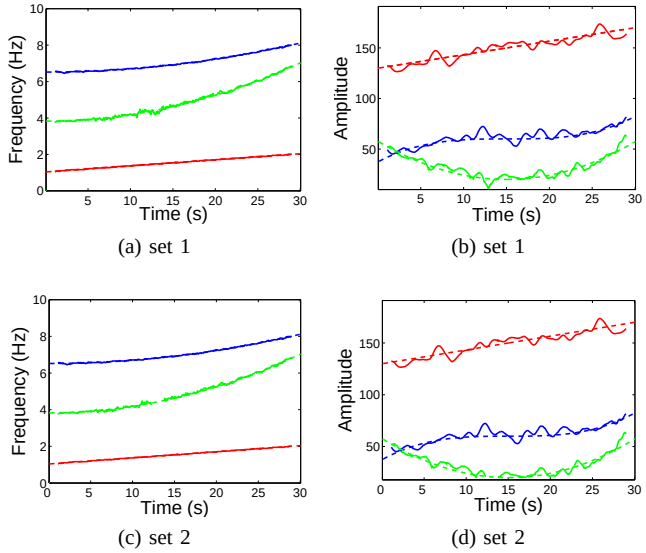


Fig. 12. Capongram of the simulated signal with the three components defined in Fig.10, generated at the global SNR of 15 dB.

Fig.13 and 14 show the results of the proposed method using the following two parameter sets that differ in $\mathcal{PFA}_d$.

set 1: $\mathcal{PFA}_e = 0.3$, $\mathcal{PFA}_d = 0.3$, $\delta_f = 2\delta_a$,$d_0 = 0.5$, $d_1 = 0.6$.

set 2: $\mathcal{PFA}_e = 0.3$, $\mathcal{PFA}_d = 0.1$, $\delta_f = 2\delta_a$,$d_0 = 0.5$, $d_1 = 0.6$.



(a) set 1 (b) set 1

(c) set 2 (d) set 2

Fig. 13. Sensitivity of $\mathcal{PFA}_d$ illustrated by the frequency and amplitude modulations of the simulated signal of Fig.10, estimated by the proposed algorithm with *set 1* and *set 2*. Dashed line, generated modulation functions; solid line, estimated modulation functions. The components are distinguished by different colors ($k = 1$,$k = 2$,$k = 3$).

The high-power components can be easily estimated with the proposed algorithm, *i.e.* the quality of the estimation is insensitive to the parameter sets. However, the weak component $k = 2$ can only be detected with a high false-alarm probability $\mathcal{PFA}_d$.

Comparing the results obtained with *set 1* and *set 2* in Fig.13, the two high-energy components $k = 1$ and $k = 3$ are not influenced by the variation of $\mathcal{PFA}_d$, since their amplitudes are strong enough to be detected with both parameter sets. However, the weak component $k = 2$ cannot be fully detected by *set 2*, which has a lower $\mathcal{PFA}_d$. The results show that the low-energy parts of the component are identified as deaths. It is common to avoid using a very high $\mathcal{PFA}_d$ value to ensure the correctness of the peak detection. The component tracking algorithm based on the successful peak detection with the $\mathcal{PFA}_d$ value in *set 1* shows excellent robustness. The impact of the total divergence thresholds d_0 and d_1 is so trivial that the estimation results remain the same as in Fig.13(a) and (b) with d_0 and d_1 taking any values in $[0.01, 1]$. No difference can be seen from Fig.13(a) and (b) if the amplitude and frequency weighting factors δ_a and δ_f vary from $\delta_f = 5\delta_a$ to $\delta_f = \delta_a/5$.

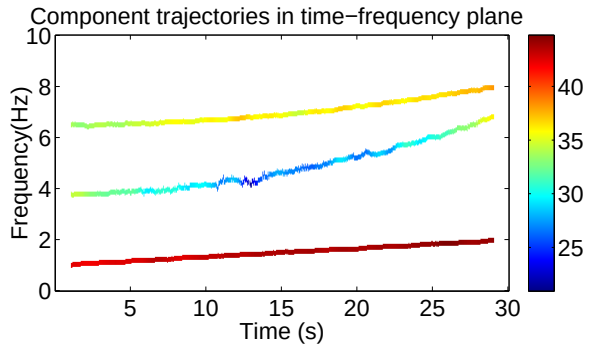


Fig. 14. Merged plot of the estimated amplitude and frequency modulations of Fig.13(a)(b) in the time-frequency plane. For each component, the frequency modulation is represented by its trajectory, while the amplitude modulation is shown by its color.

The influence of the parameter sets becomes more obvious when the noisy observation is generated at lower SNR levels. Considering *set 1* as the reference parameter set, the quality of the estimation can be studied by changing the SNR level and certain values in the test set. The Mean Absolute Errors (MAEs) of $\hat{F}_1[n]$ and $\hat{A}_1[n]$ are shown in Table III to evaluate the performance of the algorithm on the component with the highest energy.

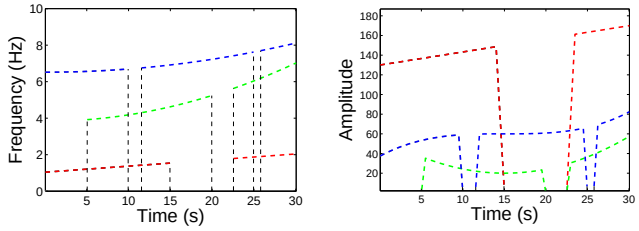
From Table III, we can infer that the performance of the algorithm is relatively robust. Under all the parameter sets, the MAEs of the frequency are of the order of 10^{-2} Hz, while the MAEs of the amplitude are also negligible. Comparing the MAEs of the frequency and of the amplitude under different SNR levels, the errors are not significantly deteriorated until the SNR reaches 0 dB, where the noise variance is equal to the average signal power. An exceptional high amplitude MAE at $\mathcal{PFA}_d = 0.1$ and low SNR level 0 dB is caused by the low false-alarm probability of detection $\mathcal{PFA}_d$, which tends to classify low-amplitude peaks as noise. For applications on signals at low SNR, a high false-alarm probability of detection is preferred at the risk of increasing false alarms, but the false peaks can be excluded at the component-tracking phase with a discrimination rule that is correctly chosen.

TABLE III
MAEs OF AMPLITUDE AND FREQUENCY $\hat{F}_1[n]$, $\hat{A}_1[n]$ AVERAGED OVER 50 NOISE REALIZATIONS, CALCULATED USING PARAMETER SET 1 AND OTHER SETS WITH DIFFERENT VALUES OF $\mathcal{PFA}_d$, $\frac{\delta_f}{\delta_a}$, d_0 OR d_1.

	Parameter sets		SNR (dB) 15	10	5	0
MAE of $\hat{A}_1[n]$	set 1		1.02	1.42	1.93	4.34
	$\mathcal{PFA}_d$	0.1	0.87	1.04	1.74	9.80
		0.5	0.88	1.22	2.19	3.83
	$\frac{\delta_f}{\delta_a}$	4	0.89	1.22	1.62	4.18
		1	0.89	1.26	2.03	3.43
	d_0	0.2	0.99	1.24	2.05	4.02
		1	0.97	1.22	1.68	3.39
	d_1	0.5	0.91	1.36	1.62	3.21
		0	0.96	1.07	1.94	3.26
MAE of $\hat{F}_1[n]$ in 10^{-2} Hz	set 1		3.31	6.92	9.56	18.96
	$\mathcal{PFA}_d$	0.1	2.39	5.42	11.07	20.37
		0.5	2.36	5.32	9.42	19.22
	$\frac{\delta_f}{\delta_a}$	4	3.30	5.15	7.61	16.46
		1	3.55	5.51	10.80	17.97
	d_0	0.2	3.60	5.65	11.45	22.62
		1	3.37	6.17	10.53	18.32
	d_1	0.5	4.06	5.45	8.42	18.02
		0	3.22	5.66	8.00	16.31

B. On multicomponent signals with births and deaths

In this section, the identification of births and deaths will be discussed. The localization of births and deaths is equivalent to the identification of the borders of the low-power regions from the time-frequency distribution. Starting from the modulation functions defined in Fig.10, several time intervals with births and deaths are added to all the components, as shown in Fig.15 and Fig.16(b). A new simulated signal is generated with number of samples $N = 600$, sampling frequency $F_s = 20$ Hz, and SNR at 15 dB. The first component $k = 1$ is dead at 15 s - 30 s ($n = 300, \ldots, 450$); the second component is born at 5 s ($n = 100$), and is dead at 20 s - 30 s ($n = 400, \ldots, 450$); the third component is dead at 10 s - 11.5 s ($n = 200, \ldots, 230$) and at 25 s - 27.5 s ($n = 500, \ldots, 515$). The Capongram of the signal is presented in Fig.16(a) besides the time-frequency plot of the estimation results.



(a) (b)

Fig. 15. Simulated frequency and amplitude modulations of a three-component simulated signal with births and deaths, where the components are distinguished by different colors ($k = 1$,$k = 2$,$k = 3$). (a) Frequency modulations $F_k[n]$. (b) Amplitude modulations $A_k[n]$.

The estimation of the modulation functions by the proposed algorithm with *set 1* is shown in Fig.17 and 18.

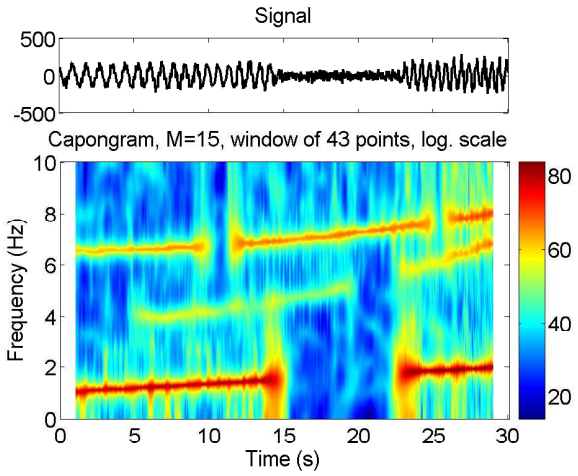
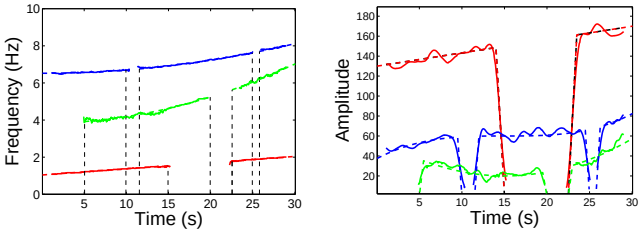


Fig. 16. (a) Capongram of the simulated signal with three components defined in Fig.15, generated at the SNR of 15 dB.

The estimation of the modulations are close to the simulated ones, whereas the instants of birth or death are detected with lags due to the time-averaging effect attributed to the sliding time window w_{L_t}. For births and deaths of long time duration, this effect might be negligible. However, the estimation might be distorted if births and deaths are short, as for $k = 3$, 10 s - 11.5 s ($n = 200, \ldots, 230$). The death of the component is estimated in such cases as a smoother amplitude attenuation than the true variation.



(a) (b)

Fig. 17. Frequency and amplitude modulations of the simulated signal with three components defined in Fig.15 at a global SNR of 15 dB, estimated by the proposed algorithm using *set 1*. The components are distinguished by different colors ($k = 1$,$k = 2$,$k = 3$). Dashed line, generated modulation functions; solid line, estimated modulation functions. (a) Frequency modulations $F_k[n]$. (b) Amplitude modulations $A_k[n]$.

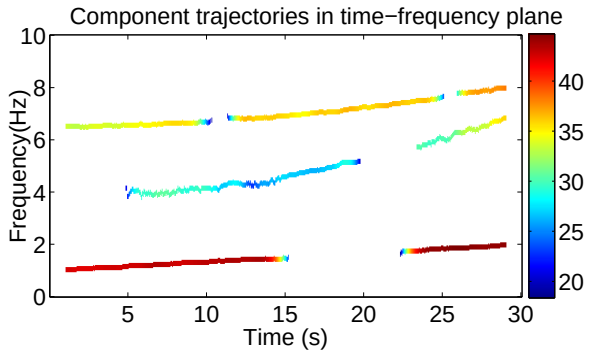


Fig. 18. Merged plot of the estimated amplitude and frequency modulations of Fig.17 in the time-frequency plane. For each component, the frequency modulation is represented by its trajectory while the amplitude modulation is shown by its color.

Comparing the Capongram of the simulated signal and the the estimation results in Fig 18, all the components are reconstructed with correct amplitude and frequency values, while the estimation of the weak component $k = 2$ is less accurate than the others because it is easier to be disturbed by the noise. The births and the deaths of all the components are identified, with estimation lags smaller than 10 points. Since the lags are affected by the length of the time window L_t, this effect cannot be avoided, but it can be alleviated by choosing a shorter window length L_t to estimate the births and the deaths accurately.

IV. Application to real-world signals

In this section, we present the application of the proposed algorithm in three real-world contexts. In Section IV-A, the algorithm is applied to vibrations of buildings, for tracking the vibratory variations under ambiant excitations or seismic ones. In section IV-B, the algorithm is tested on a canary song and in Section IV-C on a piece of oboe music.

In this section, the Capongram and the spectrogram are used according to the requirements of the application. The spectrogram is faster to compute than the Capongram, since the Fast Fourier Transform (FFT) in the calculation of the spectrogram is much faster than the calculation of the autocorrelation matrix in Capongram; secondly a spectral estimation is performed for every window in the spectrogram, while such estimation has to be done for each grid of the time-frequency plane in the Capongram. Therefore, in cases where the input signal has a high number of points, or the number of grids of the time-frequency plane is significant, it is preferable to choose spectrogram in favor of its simplicity and efficiency in the computation.

Therefore, in cases such as the oboe signal where the signal contains a sufficient number of points so that the Fourier-based method can provide a sufficient resolution in the local spectrum, the spectrogram are used. If the signals are short, or if the calculation time is not an important issue, the Capongram is used to benefit from its better performance in resolution and side-lobe reduction, as in the cases of the seismic signals. The signal of the canary song is processed using both time-frequency distributions to show a comparative study.

The computation time of the proposed algorithm which depends on the total number of FFT/Capon points of the time-frequency distribution (number of FFT/Capon points per local spectrum $\times$ number of local spectra), also on the number of iterations of the multi-pass filtering and on the number of components which defines the size of the total divergence matrix. By the way, the computation load of the Capongram is not significantly high with respect to the other steps of the proposed method.

On a computer running Matlab 2011b and Windows 7 64-bit version with an intel core i7-3540m processor and 16 GB of RAM, the execution of a non-optimized program takes around 2 μs per point per multi-pass filtering iteration. So the analysis of the oboe music in Section IV-C takes about 120 seconds in total.

A. Civil structure analysis

The assessment of the rigidity of a building is often carried out using vibration recordings from which the modal parameters of the building can be estimated. The investigation of the modal parameters and their variation allows the surveying of the damage brought to the structure, and also its characterization. The estimation of the amplitude and frequency modulations of the signal model, see *eq.* (1) is indirectly related to the physical model from which the modal parameters can be obtained.

According to the excitation sources, the vibration of a building can be categorized into two types, ambient vibrations and earthquakes [31]. In the case of ambient vibrations, the building is excited by weak and relatively stationary ground motion. With an earthquake occurrence, the ground excitation shows a sudden burst of power. Both types of vibrations are nonstationary, but the nonstationarity is stronger for an earthquake (see Fig.1) than for an ambient vibration (see Fig.19).

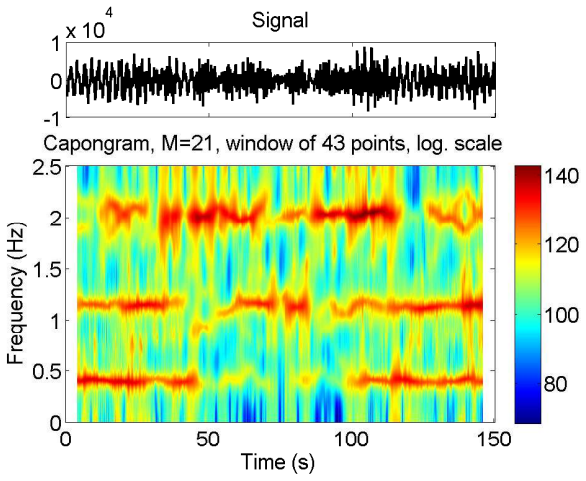
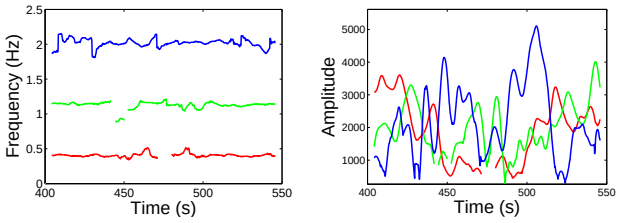


Fig. 19. Capongram of the ambient vibration recorded at the top of Taipo Tower in Taipei, Taiwan, from 400 s to 550 s, resampled at 5 Hz.

The signals processed in this section were provided by the

RISKNAT program (project URBASIS ANR-09-RISK-009, French National Research Agency).

In Fig. 20 and 21, the application of the proposed method on an ambient vibration yields three components that vary within narrow frequency bands. Using *set 1*, the total divergence values are calculated with $\delta_f = 2\delta_a$, for which the component tracking is more tolerant against the amplitude variation in Fig.21. The algorithm can track the components of drastic amplitude variations, whereas for seismic signals, as shown in Fig.22 and 23, the algorithm has to handle a different nonstationarity pattern.



(a) Frequency modulations (b) Amplitude modulations

Fig. 20. Frequency and amplitude modulations of the ambient vibration signal of Fig.19 estimated using the proposed algorithm with *set 1*. The components are distinguished by different colors ($k = 1$,$k = 2$,$k = 3$).

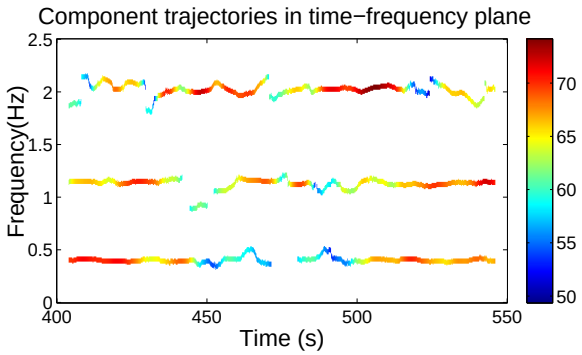


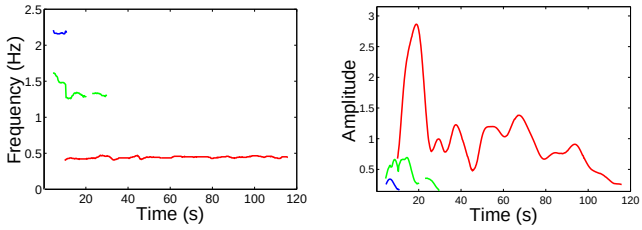
Fig. 21. Merged plot of the estimated amplitude and frequency modulations of Fig.20 in the time-frequency plane. For each component, the frequency modulation is represented by its trajectory while the amplitude modulation is shown by its color.

In the seismic recording, the power is not evenly distributed among the components, but is concentrated on a low-frequency component that is active during the entire observation. The other components are activated when the earthquake occurs, and then perish rapidly. Fig.22 shows a good estimation of these details.

B. Canary Song

The canary song has a very rich spectral content with multiple components that show smoothly varying amplitude and frequency, as illustrated in Fig.24. Similarly to many vocal signals, the richness of the components and the instaneous signal power varies drastically.

In Fig. 24, the background is cleaner on the Capongram thanks to the use if the minimum variance spectral estimator which greatly diminishes the noise. On the contrary, the spectrogram shows a more significant noise background, however



(a) Frequency modulations (b) Amplitude modulations

Fig. 22. Frequency and amplitude modulations of the seismic signal of Fig.23, estimated from the Capongram of Fig.1(a) using the proposed algorithm with *set 1*. The components are distinguished by different colors ($k = 1$,$k = 2$,$k = 3$).

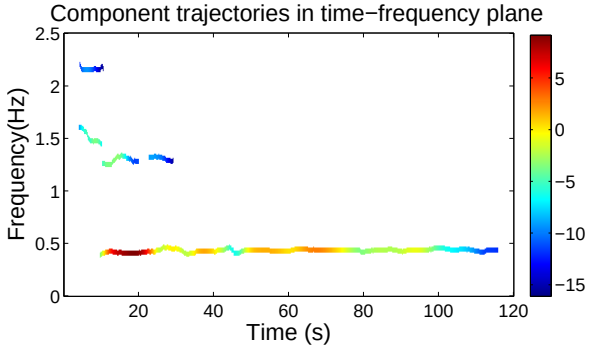


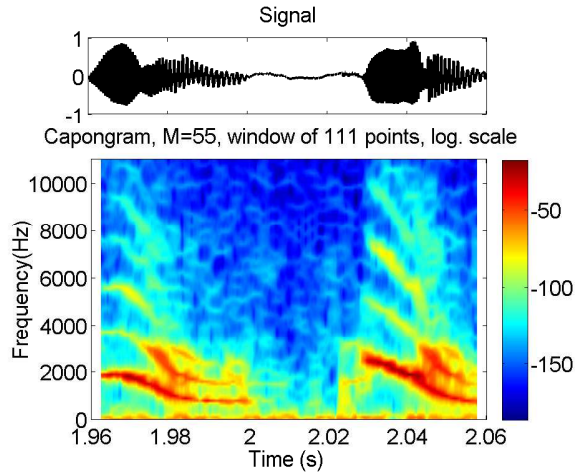
Fig. 23. Merged plot of the estimated amplitude and frequency modulations of Fig.22 in the time-frequency plane. For each component, the frequency modulation is represented by its trajectory while the amplitude modulation is shown by its color.

it preserves more detail of high frequency components which are low in amplitude.

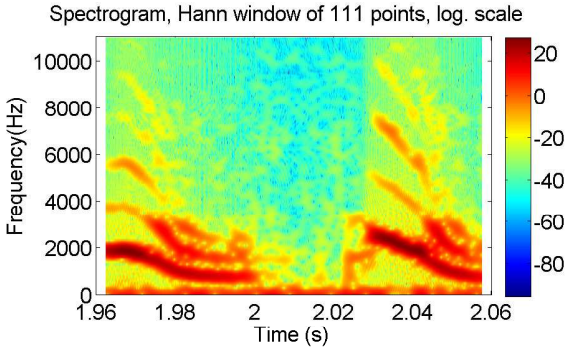
The frequency trajectories are estimated based on both the the Capongram and the spectrogram to demonstrate the adaptability of the proposed method to different time-frequency distributions chosen by the users. On the Capongram, in order to detect the high frequency components, the following parameters are chosen: $\mathcal{PFA}_e = 0.1$, $\mathcal{PFA}_d = 0.1$, $\delta_f = 2\delta_a$,$d_0 = 1$, $d_1 = 1$. While on the spectrogram, $\mathcal{PFA}_e = 0.05$, $\mathcal{PFA}_d = 0.05$ are sufficient for the detection. The results of Fig. 25 show that the five components involved are well estimated using both time-frequency distributions.

In Fig. 25, the estimation on both time-frequency distributions shows rather similar results. However some subtle differences can be noticed. At low frequency range where most of the signal power is concentrated, the modulation functions tracked on the Capongram are more detailed and less influenced by the noise. At high frequency range where the components are more difficult to be detected on the Capongram than on the spectrogram, the components tracked on the Capongram are as influenced by the noise as on the spectrogram, since a higher false alarm probability has to be set for the Capongram to detect the components.

The song is generated by air-flow-induced vibrations of the structures in the vocal organ of the canary called the syrinx. The syringeal muscles control the opening of the syrinx and therefore the frequency in the signal. The frequency trajectories decrease exponentially and become close to each

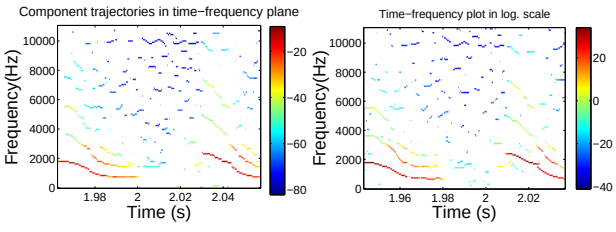


(a) Capongram of the canary song



(b) Spectrogram of the canary song

Fig. 24. Capongram and spectrogram of a canary song of 0.1 s, resampled at 22050 Hz, both are calculated using a window of 111 points.

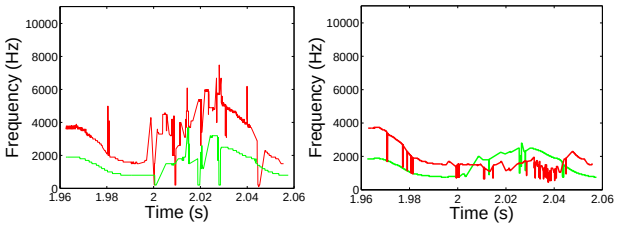


(a) (b)

Fig. 25. Merged plot of the amplitude and frequency modulations estimated respectively from the Capongram Fig. 24(a) and the spectrogram Fig. 24(b). For each component, the frequency modulation is represented by its trajectory while the amplitude modulation is shown by its color. (a) Estimated amplitude and frequency modulation functions based on the Capongram using $\mathcal{PFA}_e = 0.1$, $\mathcal{PFA}_d = 0.1$, $\delta_f = 2\delta_a$,$d_0 = 1$, $d_1 = 1$. (b) Estimated amplitude and frequency modulation functions based on the spectrogram using $\mathcal{PFA}_e = 0.05$, $\mathcal{PFA}_d = 0.05$, $\delta_f = 2\delta_a$,$d_0 = 1$, $d_1 = 1$.

other on the time intervals [1.96 s, 2 s] and [2.03 s, 2.06 s]. The components of 1.7 and 2.7 kHz corresponds to a tuned filter of the trachea and the beak.

As Fig. 26 shows, thanks to the nature of the Capon's spectral estimator, the modulation functions tracked on the Capongram are generally more continuous than on the spectrogram. The mode mixing is also rarely seen in the tracking of the powerful components. It helps to establish a more reasonable interpretation of the estimation results.



(a) from Capongram (b) from spectrogram

Fig. 26. Frequency modulations of the two most powerful components of the canary song of Fig. 25, estimated respectively from the Capongram (a) and the spectrogram (b) using the proposed algorithm. The components are distinguished by different colors ($k = 1$,$k = 2$).

In practice, the choice of the time-frequency distribution is a matter of compromise between the desired performance and the computation time.

C. Oboe music

Since the proposed method can be implemented whatever the signal type and the number of components, an oboe recording is used as an example to demonstrate its applicability on a real-world signal with a high number of components. The signal under study is illustrated in Fig.27. By the way, the proposed algorithm is applied purely on the spectrogram to demonstrate its applicability despite the choice of the time-frequency distribution.

As the spectrogram of Fig.27 shows, the signal is composed of a series of sounds, between which the power of the components dies out. During each sound, new components can also emerge and disappear. In this example, both the frequency and the amplitude of the components are estimated using the spectrogram of the signal sampled at 16000 Hz. The window length and the overlapping of the spectrogram are calculated from a required frequency resolution, which is 18 Hz.

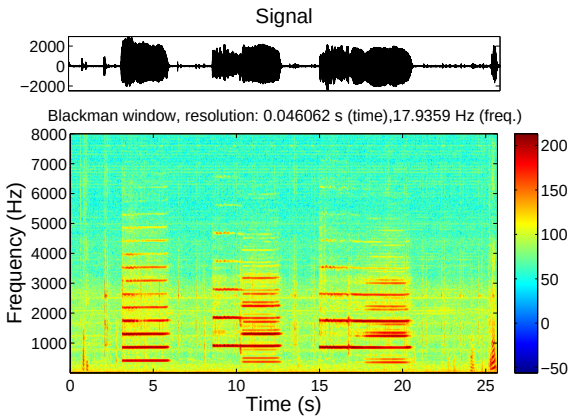
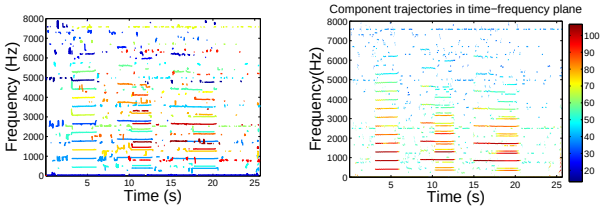


Fig. 27. A piece of oboe recording during 25.7 s, performed by Cyrille Hory and produced by Gérard Lejeune, sampled at 16000 Hz. The spectrogram is calculated using a Blackman window of 737 points (frequency resolution: 18 Hz, time resolution 46 ms) and an overlapping of 50% .

The number of components is to be estimated by applying the proposed component detection and tracking method on the spectrogram of Fig. 27. Over each local spectrum, the peaks

are detected with the false alarm probabilities $\mathcal{PFA}_d = 0.001$ and $\mathcal{PFA}_e = 0.001$ and with the multi-pass filter band L_f as equivalent to 3.5 times the width of the main lobe of the Blackman window. The configurations concerning the component tracking are the same as in *set 1*. Fig.28 shows the instantaneous frequency and amplitude of the 39 components extracted by the algorithm we propose.

In Fig. 28(b), only 1 component is tracked at the beginning, the others are detected at their own instants of activation. 18 components belong to the same harmonic family whose fundamental frequency is 440 Hz and the highest harmonic order is 17. This harmonic family is related to the key A above middle C. In fact, the player kept playing the standard concert pitch (440 Hz), this explains why the harmonic A with fundamental frequency of 440 Hz is dominant in the signal and also in the tracked components. Other components are identified as the harmonic families of other pitches which exist only within certain periods of time. For many harmonics, the trembling of the oboe tone, indicated as slight frequency fluctuations, is clearly visible and is correctly tracked by the proposed method.



(a) (b)

Fig. 28. 39 extracted components of the music signal of Fig.27 estimated from the spectrogram of Fig.27 by the proposed algorithm with $\mathcal{PFA}_d = 0.001$ and $\mathcal{PFA}_e = 0.001$, $L_f = 31$, and the configuration $\delta_f = 2\delta_a$, $d_0 = 0.05$, $d_1 = 0$ for the component tracking. (a) Estimated frequency modulations $\hat{F}_k[n]$where k is distinguished by different colors. (b) Merged plot of the estimated amplitude and frequency modulations. For each component, the frequency modulation is represented by its trajectory while the amplitude modulation is shown by its color.

Besides, a weak component which remains alive at a very low frequency (45 Hz) may correspond to the noise of the electric parts of the recording devices. There are also sudden and wide-band energy bursts which are not identified as components since the peak detection method is not adapted to the wide-band case. Since the background is not purely silent, some spurious components are detected when all the main components die out.

The estimation is robust when the false alarm probabilities are in the range of $[10^{-2}, 10^{-4}]$, $1 \leq \frac{\delta_f}{\delta_a} \leq 5$ and $0.03 \leq d_0 \leq 0.2$. Nevertheless, the estimation is degraded if the frequency resolution of the spectrogram is greater than 24 Hz, because the trajectories of the pitches may be gathered in the time-frequency plane.

V. Conclusion

In this paper, we are interested in the component extraction of non-stationary signals with births and deaths. We proposed an original method to detect the peaks of a time-frequency distribution and to form the components by tracking the detected peaks in time. The method starts from a time-frequency distribution of the signal. At each local spectrum, a peak detection method is applied to extract the peaks from the noise spectrum, which is estimated by a multipass filtering method. Sequentially in time, the detected peaks are tracked with previously formed components by evaluating the total divergence matrix, where the minimum values indicate the peak-component pairs to be selected. The elements of the total divergence matrix indicate the divergence of frequency and amplitude between a detected peak and a formed component.

The performance analysis assessed on simulated signals shows that the proposed method is accurate and robust despite low SNR levels and across different parameter sets. The number of components is estimated and the components are tracked automatically. Due to two adjustable false-alarm probabilities, the peak detection from the noise spectrum is relatively flexible. In the component-tracking step, a peak-component classifier can automatically track peaks to build components. The births and the deaths of components can be automatically identified from the existence or the absence of peaks. The total divergence matrix proposed in this paper is a useful distance metric that helps to select unique peak-component pairs.

The proposed tracking method takes into account not only the frequency, but also the amplitude. By the way it permits the tracking of a trajectory with discontinuation, which is another novelty of the approach compared to the others.

As a general-purpose method, the applicability of the proposed method is illustrated on a civil structure evaluation, a bird song analysis and an oboe music processing.

As the proposed method is based on the time-frequency distribution of the signal, high frequency and time resolution cannot be achieved at the same time. The margins of the births and the deaths of the components are estimated with time lags due to time averaging effects that cannot be avoided by the proposed method. To solve these problems, work is in progress in order to propose a local signal model where the amplitude and the frequency are approximated by low-order polynomials and the births and the deaths of the components are modeled by Heaviside functions. The model parameters will be estimated using a maximum likelihood strategy. Such a method will no longer be based on an *a priori* time-frequency estimation and thus will require no analysis window causing the time-lag effect. In this way the estimation of the modulations will be carried out at a very high precision and resolution as it is the case for the parametric approaches. The method proposed in the present paper will serve as an initialization step to provide a preliminary estimation that can be refined by the parametric approach. The interest of the non-parametric approach proposed in this paper lies in a gain of computation time and in the capability of identifying a non-limited number of signal components.

Appendix
Capongram and Spectrogram

To demonstrate the applicability of the proposed algorithm on both Fourier-based and non-Fourier time-frequency distributions, the two time-frequency distributions considered in this

paper: the spectrogram and the Capongram. The spectrogram of a discrete signal $y[n]$ is a commonly-used tool in time-frequency analysis,

$$\mathcal{TFD}_y[n,f] = \left| \sum_{l=-L_t/2}^{L_t/2} w^*_{L_t}[l-n]y[l]e^{-\frac{\pi j f}{L_t+1}} \right|^2. \quad (27)$$

with $w[n]$ a windows function of L_t+1 points. Whereas for the Capongram, a piecewise spectrum is estimated using a non-Fourier estimator, namely the Capon's method [32], also known as the Minimum Variance Spectral Estimator (MVSE). Instead of using a Fourier kernel, the Capon's method designs a data-adaptive filter $[h_{L_t}[-L_t/2],\ldots,h_{L_t}[L_t/2]]$ which minimizes the total energy output

$$J_{out} = \left| \sum_{l=-L_t/2}^{L_t/2} h^*_{L_t}[l-n]y[l]e^{-\frac{\pi j f}{L_t+1}} \right|^2, \quad (28)$$

subject to

$$J_{penal} = \left| \sum_{l=-L_t/2}^{L_t/2} h_{L_t}[l]e^{-\frac{\pi j f}{L_t+1}} \right|^2 = 1. \quad (29)$$

Thus, the obtained filter parameters

$$\mathbf{h}_{L_t} = [h_{L_t}[-L_t/2],\ldots,h_{L_t}[L_t/2]] \quad (30)$$

should preserve the energy at frequency f while it rejects the energy at other frequencies. This constraint optimization problem can be solved by the Lagrange multiplier method. The cost function

$$J = J_{out} - \lambda\left(J_{penal} - 1\right) \quad (31)$$

is minimized when

$$\mathbf{h}_{L_t} = \frac{\mathbf{R}_M^{-1}\mathbf{E}_f}{\mathbf{E}_f^H\mathbf{R}_M^{-1}\mathbf{E}_f}, \qquad \lambda = \frac{1}{\mathbf{E}_f^H\mathbf{R}_M^{-1}\mathbf{E}_f} \quad (32)$$

where $\mathbf{E}_f = [1, e^{\frac{2\pi j f}{F_s}},\ldots,e^{\frac{2\pi j(M-1)f}{F_s}}]$ with F_s being the sampling frequency of the signal, M is the order of the Capon's filter, and $\mathbf{R}_M$ is the $M\times M$ auto-correlation matrix of the windowed signal $y[n-\frac{L_t}{2}],\ldots,y[n+\frac{L_t}{2}]$.

Substituting $\mathbf{h}_{L_t}$ in (28), the estimated power spectrum of the data-adaptive filter at frequency f is

$$\mathcal{TFD}_y[n,f] = \frac{1}{\mathbf{E}_f^H\mathbf{R}_M^{-1}\mathbf{E}_f}. \quad (33)$$

The Capon's method is known to have higher resolution than the Fourier-based spectrum estimators [27] [28]. Due to the energy minimization of the filter design, the side-lobes of the frequency response of the filter are significantly reduced compared to the Fourier-based methods using a smoothing window.

REFERENCES

[1] F. Millioz and N. Martin, "Time-Frequency Segmentation for Engine Speed Monitoring," in *ICSV13*, Vienna, Austria, July 2-6 2006.

[2] C. Michel and P. Guegen, "Time-frequency analysis of small frequency variations in civil engineering structures under weak and strong motion using reassignment method," in *Structural Health Monitoring*, 2010.

[3] V. Rai and A. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform," *Mechanical Systems and Signal Processing*, vol. 21, pp. 2607–2615, 2007.

[4] X. Qiuzhen and B. Shankara Reddy, "Recognition of high-resolution ECGs by time-frequency representation," in *Proceedings of the 16th Annual International Conference of the IEEE*, 1994, pp. 1248–1249.

[5] S. Meignen, T. Oberlin, and S. McLaughlin, "A New Algorithm for Multicomponent Signals Analysis Based on SynchroSqueezing: With an Application to Signal Sampling and Denoising," *IEEE Transactions on Signal Processing*, vol. 60, pp. 5787–5798, 2012.

[6] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the royal society A*, vol. 459, no. 2037, pp. 2317–2345, 1998.

[7] X. Serra and Smith J., "Spectral Modeling Synthesis A Sound Analysis/Synthesis Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, vol. 14, pp. 12–24, 1990.

[8] Chinmay Pendharkar, "Auralization of road vehicles using spectral modeling synthesis," Ph.D. dissertation, Chalmers University of Technology, 2012.

[9] J. Laroche, Y. Stylianou, and E. Moulines, "HNM: a simple, efficient harmonic+noise model for speech," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1993, pp. 169–172.

[10] Y. Pantazis, O. Rosec, and Y. Stylianou, "Adaptive AM-FM signal secomposition with application to speech analysis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 290–300, 2011.

[11] R. McAulay, "Speech analysis/Synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, Issue 4, pp. 744–754, 1986.

[12] P. Depalle, G. Garcia, and X. Rodet, "analysis of sound for additive synthesis: Tracking of partials using hidden markov models,," in *Proc. ICMC*, 1993.

[13] S. Barbarossa, "Analysis of multicomponent LFM signals by a combined Wigner-Hough transform," *IEEE Transactions on Signal Processing*, vol. 43, Issue 6, pp. 1511–1515, 1995.

[14] L. Cirillo, A. Zoubir, and M. Amin, "Parameter Estimation for Locally Linear FM Signals Using a Time-Frequency Hough Transform," *IEEE Transactions on Signal Processing*, vol. 56, Issue 9, pp. 4162–4175, 2008.

[15] V. Katkovnik, "Nonparametric estimation of instantaneous frequency," *IEEE Transactions on Information Theo*, vol. 30, pp. 183–189, 1997.

[16] V. Katkovnik and L. Stankovic, "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length," *IEEE Transactions on Signal Processing*, vol. 46, pp. 2315–2325, 1998.

[17] I. Djurovic and L. Stankovic, "An algorithm for the Wigner distribution based instantaneous frequency estimation in a high noise environment," *Signal*, vol. 84, Issue 3, pp. 631–643, 2004.

[18] L. Rankine, M. Mesbaha, and B. Boashash, "IF estimation for multicomponent signals using image processing techniques in the timecfrequency domain," *Signal Processing*, vol. 87, Issue 6, pp. 1234–1250, 2007.

[19] D. Vakman, "On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency," *IEEE Transactions on Signal Processing*, vol. 44, Issue 6, pp. 791–797, 1996.

[20] J. R. Beltran and J. Ponce de Leon, "Estimation of the instantaneous amplitude and the instantaneous frequency of audio signals using complex wavelets," *Si*, vol. 90, Issue 12, pp. 3092–3109, 2010.

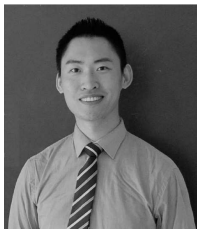
[21] K. S. Arun and R. M. Liang, "Instantaneous amplitude and frequency estimation: performance bounds and applications to source localization," *Proc. SPIE*, vol. 1566, pp. 156–166, 1991.

[22] M. Durnerin and N. Martin, "Minimum Variance Filters and Mixed Spectrum Estimation," *Signal Processing*, vol. 80, Issue 12, pp. 2597–2608, 2000.

[23] N. Martin, J. Mars, J. Martin, and C. Chorier, "A Capon's time-octave representation application in room acoustics," *IEEE Transactions on Signal Processing*, vol. 43, Issue 8, pp. 1842–1854, 1995.

[24] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[25] L. Cohen, P. Loughlin, and D. Vakman, "On an ambiguity in the definition of the amplitude and phase of a signal," *Signal Processing*, vol. 79, Issue 3, pp. 301–307, 1999.
[26] E. Bedrosian, "The analytic representation of modulated waveforms," *Proceedings of the IRE*, vol. 50, Issue 10, pp. 2071–2076, 1962.
[27] A. Kayhan, A. El-Jaroudi, and L. Chaparro, "Data-adaptive evulationary spectral estimation, ieee trans. sp," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 204–213, 1995.
[28] N. Martin, *Spectral Analysis, Parametric, Non-parametric and Advanced Methods*, F. CASTANIE, Ed. WILEY-VCH, 2011.
[29] M. Durnerin, "A strategy for interpretation in spectral analysis," Ph.D. dissertation, Institut National Polytechnique de Grenoble, France, 1992.
[30] C. Mailhes, N. Martin, K. Sahli, and G. Lejeune, "A Spectral Identiy Card," in *EUropean SIgnal Processing Conference, EUSIPCO 06,*, Florence, Italy, 2006.
[31] F. Dunand, P. Gueguen, P. Bard, J. Rodgers, and M. Celebi, "Comparison of the dynamic parameters extracted from weak, moderate and strong motion recorded in buildings," in *First European Conference on Earthquake Engineering and Seismology*, 2006.
[32] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.



Zhongyang Li Zhongyang li was born in P. R. China in 1985. He received the Bachelor and Master degree in engineering from Southeast University, Nanjing, P. R. China, in 2007 and 2010 respectively. He received the Ph. D. degree in signal processing at Gipsa-lab/Grenoble-INP, France in 2013. He is currently a post-doctoral researcher at SAIGA team, Gipsa-lab. His research interests include time-frequency analysis, spectral analysis and signal processing for fault diagnosis.

Nadine Martin Nadine Martin received his Engineering degree in Physics and Electronics from the Institute of Chemistry and Physics of Lyon, France, in 1980 and the Ph.D. degree in Signal Processing and Control from Grenoble Institute of Technology, France, in 1984. Currently, she is a senior researcher at the CNRS, Nat. Centre of Scientific Research, in the research team SAIGA, *Signal and Automatic for surveillance, diagnostic and biomechanics*, a team within GIPSA-lab, *Grenoble Image Speech Signal & Automatic*, Grenoble, France. In the signal-processing domain, her research interests are the analysis, detection and modeling of non-stationary signals in time and in time-frequency domain with a particular interest in the condition monitoring of complex systems. N. Martin is the author of more than 150 papers, 2 patents, 7 book chapters, and the co-editor of the book Time-frequency decision. She is a member of the Management Board of ISCM, Int. Society for Condition Monitoring, Vice-Editor-in-Chief of IJCM, Int. Journal of Condition Monitoring and member of the Int. ICNDT/ISCM working group of the Int. Committee of Nondestructive testing.