Probabilistic Time-Frequency Source-Filter Decomposition of Non-Stationary Signals

Roland Badeau, Mark. D. Plumbley

PROBABILISTIC TIME-FREQUENCY SOURCE-FILTER DECOMPOSITION OF NON-STATIONARY SIGNALS

Roland Badeau

Institut Mines-Telecom
Telecom ParisTech
CNRS LTCI

Mark D. Plumbley

Centre for Digital Music, School of
Electronic Engineering & Computer Science
Queen Mary University of London

ABSTRACT

Probabilistic modelling of non-stationary signals in the time-frequency (TF) domain has been an active research topic recently. Various models have been proposed, notably in the nonnegative matrix factorization (NMF) literature. In this paper, we propose a new TF probabilistic model that can represent a variety of stationary and non-stationary signals, such as autoregressive moving average (ARMA) processes, uncorrelated noise, damped sinusoids, and transient signals. This model also generalizes and improves both the Itakura-Saito (IS)-NMF and high resolution (HR)-NMF models.

***Index Terms*—** Probabilistic modelling, Non-stationary processes, Time-frequency analysis, Source-filter models, Nonnegative matrix factorisation.

1. INTRODUCTION

In the literature, several probabilistic models involving latent components have been proposed for modelling TF representations of audio signals such as spectrograms. Such models include NMF with additive Gaussian noise [1], probabilistic latent component analysis (PLCA) [2], NMF as a sum of Poisson components [3], and NMF as a sum of Gaussian components [4]. Although they have already proved successful in a number of audio applications such as source separation [2, 3] and multipitch estimation [4], most of these models still lack of consistency in some respects. Firstly, they focus on a magnitude or power TF representation, and simply ignore the phase information. As a consequence, reconstructing a consistent TF representation, including the phase field, often proves to be difficult [5, 6]. Secondly, these models generally focus on the spectral and temporal dynamics, and assume that all time-frequency bins are independent. This assumption is not consistent with the existence of signal dynamics, because spectral dynamics always induces temporal dependencies, and temporal dynamics always induces spectral dependencies. In addition, further dependencies in the TF domain may be induced by the TF transform, due to spectral and temporal overlap between TF bins. In this paper, we address this problem by introducing a new probabilistic model called *probabilistic time-frequency source-filter decomposition* (PTFSFD), which aims to take both phases within TF bins and correlations between TF bins into account. It consists of a sum of source-filter models, where each component is obtained (in the original time domain) by successively applying a multiplication and a convolution to a white noise.

The paper is structured as follows. The filter bank used to compute the TF representation is introduced in Section 2. We then show in Sections 3 and 4 how convolutions and multiplications in the original time domain can be accurately implemented in the TF domain. The PTFSFD model is introduced in Section 5, and some examples are provided in Section 6. Finally, conclusions are drawn in Section 7.

2. DEFINITION OF THE FILTER BANK

In order to properly define the PTFSFD model in the TF domain, we need to accurately implement convolutions and multiplications. In the literature, the Short Time Fourier Transform (STFT) [7] is often considered as a convenient TF transform, because under some smoothness assumptions it allows the approximation of convolution by multiplying each column of an STFT by the same spectrum, and of multiplication by multiplying each row of an STFT by the same waveform. However, STFT is an oversampled filter bank, and it transforms a real random process into a complex TF random field, whose distribution is generally approximated as circularly symmetric[1]. Instead, we propose to use a critically sampled perfect reconstruction (PR) cosine-modulated filter bank [7], which has several desired properties. Firstly, uniform sampling in both time and frequency will allow an accurate implementation of convolutions and multiplications in the TF domain (*cf.* Sections 3 and 4). Secondly, real TF distributions will permit us to avoid the need for any circularly-symmetric assumption. The combination of PR, paraconjugate[2] analysis/synthesis filters and critical sampling will al-

[1] A complex scalar random variable Z is circularly symmetric if $\forall \Psi \in \mathbb{R}$, the random variables Z and $Ze^{i\Psi}$ have the same probability distribution.

[2] The paraconjugation of a discrete linear filter consists of the time reversal and complex conjugation of its impulse response.

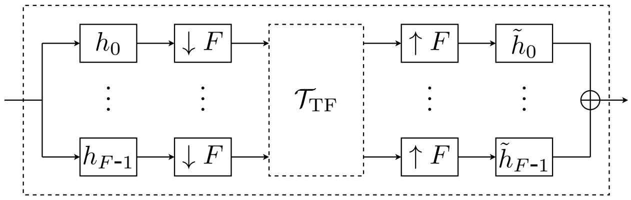

Fig. 1. Analysis/synthesis filter bank

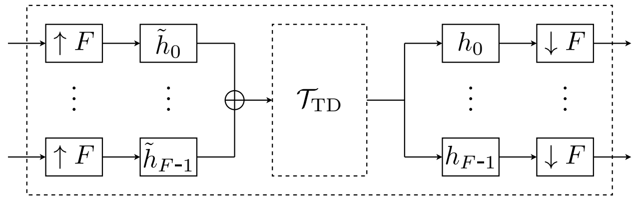

Fig. 2. Synthesis/analysis filter bank

low us to ensure the filter bank does not introduce any spurious correlation artefacts, concept that will be referred to as *preservation of whiteness* (PW), and explained in Section 5.1.

We thus consider in this paper a critically sampled cosine-modulated filter bank [7], which transforms an input signal $x(n) \in \mathbb{R}$ in the original time domain $n \in \mathbb{Z}$ into a 2D-array $x(f,t) \in \mathbb{R}$ in the TF domain $(f,t) \in [0 \dots F-1] \times \mathbb{Z}$. More precisely, $x(f,t)$ is defined as $x(f,t) = (h_f * x)(Ft)$, where $*$ denotes the standard convolution of non-periodic series over $\mathbb{Z}$, $h_f(n) = h(n) \cos\left(\frac{\pi}{F}(f+\frac{1}{2})(n+\phi)\right)$ is a real-valued analysis filter, F is an odd number so that $\phi = \frac{F+1}{2}$ is a whole number, and $h(n)$ is a prototype window of support $[0 \dots N-1]$ with $N = LF$ and $L \in \mathbb{N}$, whose frequency passband is $[-\frac{1}{2F}, \frac{1}{2F}]$ (modulo 1). The synthesis filters are defined as $\tilde{h}_f(n) = h_f(N-n)$. This analysis/synthesis filter bank, represented in Figure 1 ($\mathcal{T}_{\text{TF}}$ is defined as the TF-domain identity within Section 2), is designed so as to guarantee PR, which requires that L be even [7]. This means that the output, defined as $x'(n) = \sum_f \sum_t \tilde{h}_f(n-Ft)x(f,t)$, satisfies $x'(n) = x(n-N)$. The same PR property also guarantees that if $x(f,t)$ is the input of the synthesis/analysis filter bank represented in Figure 2 ($\mathcal{T}_{\text{TD}}$ is defined as the time-domain identity within Section 2), then the output is $x'(f,t) = x(f,t-L)$. Let $H(\nu) = \sum_{n\in\mathbb{Z}} h(n)e^{-2i\pi\nu n}$ (with an upper case letter) denote the discrete time Fourier transform (DTFT) of $h(n)$ over $\nu \in \mathbb{R}$. Considering that the time supports of $h(Ft_1 - n)$ and $h(Ft_2 - n)$ do not overlap provided that $|t_1 - t_2| \geq L$, we similarly define a whole number K, such that the overlap between the frequency supports of $H_{f_1}(\nu)$ and $H_{f_2}(\nu)$ can be neglected provided that $|f_1 - f_2| \geq K$, due to high rejection in the stopband. We note that $h_f(n)$ and $x(f,t)$ are actually defined for all frequencies $f \in \mathbb{Z}$, with the following properties:

- $\forall n, t \in \mathbb{Z}$, $h_f(n)$ and $x(f,t)$ are $2F$-periodic w.r.t. f,
- $\forall n, t \in \mathbb{Z}$, $h_f(n)$ and $x(f,t)$ are symmetric w.r.t. $F-\frac{1}{2}$: $\forall f \in \mathbb{Z}, x(f,t) = x(2F-1-f,t)$.

3. TF IMPLEMENTATION OF A CONVOLUTION

In this section, we consider a filter of impulse response $g(n) \in \mathbb{R}$ and two signals $x(n) \in \mathbb{R}$ and $y(n) \in \mathbb{R}$, such that $y(n) = (g * x)(n)$. Our purpose is to directly express $y(f,t)$ as a function of $x(f,t)$, *i.e.* to find a TF transformation $\mathcal{T}_{\text{TF}}$ in Figure 1 such that the output of the filter bank is $y(n-N)$ when its input is $x(n)$. To achieve this, we use the PR property of this critically sampled filter bank, which implies that the unique solution $\mathcal{T}_{\text{TF}}$ is that described in the larger frame in Figure 2, where the input is $x(f,t)$, the output is $y(f,t)$, and transformation $\mathcal{T}_{\text{TD}}$ is defined as the time-domain convolution by $g(n+N)$. We then obtain

$$y(f,t) = \sum_{\varphi=-F}^{F-1} \sum_{\tau\in\mathbb{Z}} c_g(f,\varphi,\tau)\, x(f-\varphi, t-\tau), \qquad (1)$$

where $\forall \varphi \in [-F \dots F-1]$, we have $\forall f \in [0 \dots F-1]$,

$$c_g(f,\varphi,\tau) = (h_f * \tilde{h}_{f-\varphi} * g)(F(\tau+L)), \qquad (2)$$

and $\forall f \in [F \dots 2F-1]$, $c_g(f,\varphi,\tau) = c_g(2F-1-f-\varphi,\tau)$, where we have used the convention $\forall \varphi \notin [0 \dots F-1]$, $h_\varphi = 0$. This definition of $c_g(f,\varphi,\tau)$ is extended to all $f \in \mathbb{Z}$ and $\varphi \in \mathbb{Z}$ by $2F$-periodicity. Note that $\forall \varphi \in [-F \dots -K] \cup [K \dots F-1]$, subbands f and $f-\varphi$ do not overlap, thus $c_g(f,\varphi,\tau)$ can be neglected. Equation (1) shows that a convolution in the original time domain is equivalent to a 2D-convolution in the TF domain, which is stationary w.r.t. time, and circular of period $2F$ but non-stationary w.r.t. frequency.

In the following definition, we propose a parametric model of $c_g(f,\varphi,\tau)$. In Appendix A we prove that this model can represent any causal and stable ARMA filter $g(n)$.

Definition 1. *The ARMA model of convolution in the TF domain is defined by* (1)*, where filters* $\{c_g(f,\varphi,\tau)\}_{\substack{0\leq f<F \\ -F\leq\varphi\leq F-1}}$ *are defined as the only stable solutions of the following equation:* $\forall f, \varphi, \tau$,

$$\textstyle\sum_{u=0}^{Q_a} a_g(f-\varphi,u) c_g(f,\varphi,\tau-u) = b_g(f,\varphi,\tau). \qquad (3)$$

In equation (3)*, the autoregressive term* $a_g(f-\varphi,u)$ *is a causal sequence of support* $[0 \dots Q_a]$ *(where* $Q_a \in \mathbb{N}$*) w.r.t.* u*, having only simple poles lying inside the unit circle, the moving average term* $b_g(f,\varphi,\tau)$ *is a sequence of support* $[-L+1 \dots -L+1+Q_b]$ *(where* $Q_b \geq 2L+Q_a-1$*) w.r.t.* τ*, and* $\forall f \in [0 \dots F-1]$*,* $\forall \varphi \in [-F \dots F-1]$*, if* $f-\varphi \notin [0 \dots F-1]$ *or* $|\varphi| > P_b$ *with* $P_b = K-1$*, then* $b_g(f,\varphi,\tau)$=0.

Remark 1. If $a_g(f,u)$ and $b_g(f,\varphi,\tau)$ are arbitrary, then there may be no filter $g(n)$ such that equation (2) holds, which means that the operation defined in Definition 1 does no longer correspond to a convolution in the original time domain. In this case, we will say that the model is *inconsistent*.

4. TF IMPLEMENTATION OF A PRODUCT

In this section, we consider a sequence $\sigma(n) \in \mathbb{R}$ and two signals $w(n) \in \mathbb{R}$ and $x(n) \in \mathbb{R}$, such that $x(n) = \sigma(n)w(n)$. As in the previous section, our purpose is to directly express $x(f,t)$ as a function of $w(f,t)$, *i.e.* to find a TF transformation $\mathcal{T}_{\text{TF}}$ in Figure 1 such that the output of the filter bank is $x(n-N)$ when its input is $w(n)$. To achieve this, we also use the PR property of the filter bank, which implies that the unique solution is that represented inside the larger frame in Figure 2, where the input is $w(f,t)$, the output is $x(f,t)$, and transformation $\mathcal{T}_{\text{TD}}$ is defined as the time-domain convolution by $\delta(n+N)$ (δ is the Kronecker symbol: $\delta(u) = 1$ if $u = 0$, $\delta(u) = 0$ otherwise) followed by the product by $\sigma(n)$. Mathematically, we obtain $x(f,t) = \sum_{n\in\mathbb{Z}} \sigma(n) h_f(Ft - n) \sum_{\varphi=0}^{F-1} \sum_{\tau\in\mathbb{Z}} h_\varphi(F\tau - n) w(\varphi,\tau)$, which is equivalent to

$$\begin{array}{rcl} x(f,t) & = & (-1)^{ft} \sum_{\varphi=-F}^{F-1} \sum_{\tau\in\mathbb{Z}} p_\sigma(t,\varphi,\tau) \\ & & \times(-1)^{(f-\varphi)(t-\tau)} w(f-\varphi, t-\tau) \end{array} \quad (4)$$

$$\text{where } p_\sigma(t,\varphi,\tau) = \frac{(-1)^{\varphi t}}{2} \sum_{m\in\mathbb{Z}} \sigma(Ft-m) h(m) \times h(m - F\tau) \cos\left(\frac{\pi}{F}\varphi(m+\phi) + \frac{\pi}{2}\tau\right). \quad (5)$$

Here, it is important to note that $\forall t \in \mathbb{Z}$,

- $\forall \tau \in \mathbb{Z}$, $p_\sigma(t,\varphi,\tau)$ is $2F$-periodic w.r.t. φ,
- if τ is even, then $p_\sigma(t,\varphi,\tau)$ is symmetric w.r.t. φ,
- if τ is odd, then $p_\sigma(t,\varphi,\tau)$ is antisymmetric w.r.t. φ,
- $\forall \tau \in \mathbb{Z}$, $p_\sigma(t,\varphi,-\tau) = (-1)^\tau p_\sigma(t+\tau,\varphi,\tau)$,
- $\forall \tau \notin [-L+1 \dots L-1]$, $p_\sigma(t,\varphi,\tau) = 0$.

Equation (4) shows that a product in the original time domain is equivalent to a 2D-convolution in the TF domain, which is circular and stationary w.r.t. frequency, but non-stationary w.r.t. time. In Definition 2, we propose a parametric model of $p_\sigma(t,\varphi,\tau)$. We prove in Appendix B that this model can represent a variety of sequences $\sigma(n)$, ranging from slowly varying sequences to linear combinations of impulses.

Definition 2. *The ARMA model of multiplication in the TF domain is defined by equation* (4)*, where filters* $\{p_\sigma(t,\varphi,\tau)\}_{\substack{t\in\mathbb{Z}\\ \tau\in\mathbb{Z}}}$ *are defined as the only* $2F$*-periodic solutions of the following equation:* $\forall t,\varphi,\tau$,

$$\sum_{\omega=-P_\alpha}^{P_\alpha} \alpha_\sigma(t-\tau,\omega) p_\sigma(t,\varphi-\omega,\tau) = \sum_{u\in\mathbb{Z}} \beta_\sigma(t,\varphi-2Fu,\tau). \quad (6)$$

In equation (6)*, the autoregressive term* $\alpha_\sigma(t-\tau,\omega)$ *is a non-periodic symmetric sequence of bounded support* $[-P_\alpha \dots P_\alpha]$ *(where* $P_\alpha < F-K$*) w.r.t.* ω*, the moving average term* $\beta_\sigma(t,\varphi,\tau)$ *is a non-periodic sequence of bounded support* $[-P_\beta \dots P_\beta]$ *(where* $K + P_\alpha \le P_\beta < F$*) w.r.t.* φ*, and*

- *if* τ *is even, then* $\beta_\sigma(t,\varphi,\tau)$ *is symmetric w.r.t.* φ*,*
- *if* τ *is odd, then* $\beta_\sigma(t,\varphi,\tau)$ *is antisymmetric w.r.t.* φ*,*
- $\forall \tau \notin [-Q_\beta \dots Q_\beta]$ *with* $Q_\beta = L-1$, $\beta_\sigma(t,\varphi,\tau) = 0$.

Remark 2. If $\alpha_\sigma(t,\omega)$ and $\beta_\sigma(t,\varphi,\tau)$ are arbitrary, then there may be no sequence $\sigma(n)$ such that equation (5) holds, which means that the operation defined in Definition 2 does not correspond to a multiplication in the original time domain. In this case, we will say that the model is *inconsistent*.

5. TF PROBABILISTIC MODELLING

The PTFSFD model introduced in Section 5.2 relies on the fundamental PW property presented in Proposition 1.

5.1. Preservation of whiteness

Proposition 1 (Preservation of whiteness)**.** *The filter bank defined in Section 2 guarantees that a stochastic process* $w(n)$ *is white noise of variance* σ^2 *if and only if its TF counterpart* $w(f,t)$ *is 2D-white noise of same variance* σ^2*.*

Proof. If $w(n)$ is white noise of variance σ^2, then $\forall f_1, f_2 \in [0 \dots F-1]$, $\forall t_1, t_2 \in \mathbb{Z}$, $\mathbb{E}(w(f_1,t_1)w(f_2,t_2)) = \sigma^2 \times \sum_{n\in\mathbb{Z}} h_{f_1}(Ft_1 - n) h_{f_2}(Ft_2 - n) = \sigma^2 \delta(f_1 - f_2)\delta(t_1 - t_2)$, where $\mathbb{E}$ denotes mathematical expectation. The second equality comes from the PR property of the synthesis/analysis filter bank (*cf.* Figure 2, where $\mathcal{T}_{\text{TD}}$ is defined as the time-domain identity). Thus $w(f,t)$ is 2D-white noise of variance σ^2. Reciprocally, if $w(f,t)$ is 2D-white noise of variance σ^2, then $\mathbb{E}(w(n_1)w(n_2)) = \sigma^2 \sum_{f=0}^{F-1} \sum_{t\in\mathbb{Z}} h_f(Ft - n_1) h_f(Ft - n_2) = \sigma^2 \delta(n_1 - n_2)$, where the second equality comes from the PR property of the analysis/synthesis filter bank (*cf.* Figure 1, where $\mathcal{T}_{\text{TF}}$ is defined as the TF-domain identity). Thus $w(n)$ is white noise of variance σ^2. $\square$

5.2. Definition of the PTFSFD model

Consider a mixture signal $y(n) = \sum_{r=0}^{R-1} y_r(n)$, where each component $y_r(n)$ follows a source-filter model: $y_r = g_r * x_r$, where $g_r(n)$ is an impulse response of the form defined in Appendix A, $x_r = \sigma_r \times w_r$, where $\sigma_r(n)$ is a sequence of the form defined in Appendix B, and $w_r(n)$ is white noise of variance 1. Then in the TF domain, $y(f,t)$ satisfies the PTFSFD model presented in Definition 3.

Definition 3. *The PTFSFD model expresses a TF representation* $y(f,t)$ *(with* $0 \le f \le F-1$*) as a sum of* R *uncorrelated components* $\{y_r(f,t)\}_{0\le r<R}$*, where*

$$y_r(f,t) = \sum_{\varphi=-F}^{F-1} \sum_{\tau\in\mathbb{Z}} c_{g_r}(f, f-\varphi, t-\tau) x_r(\varphi,\tau),$$

$$x_r(f,t) = (-1)^{ft} \sum_{\varphi=-F}^{F-1} \sum_{\tau\in\mathbb{Z}} p_{\sigma_r}(t, f-\varphi, t-\tau) \left((-1)^{\varphi\tau} w_r(\varphi,\tau)\right),$$

$w_r(f,t)$ *is 2D-white noise of variance* 1*, and* $c_{g_r}(f,\varphi,\tau)$ *and* $p_{\sigma_r}(t,\varphi,\tau)$ *are parametrised as in Definitions 1 and 2.*

6. DEMONSTRATION EXAMPLES

The PTFSFD model introduced in Definition 3 encompasses the following particular cases:

1) **ARMA process:** If $\forall t$, $p_r(t,\varphi,\tau)$=$\delta_{2F}(\varphi)\delta(\tau)$ (where $\delta_{2F}(\varphi) = \sum_{u\in\mathbb{Z}} \delta(\varphi - 2Fu)$), then $y_r(n)$ can be any ARMA process (an example is given in Figures 3-1a and 3-1b).

2) **Uncorrelated noise:** If $\forall f$, $c_{g_r}(f,\varphi,\tau) = \delta_{2F}(\varphi)\delta(\tau)$, $y_r(n)$ belongs to a parametric family of uncorrelated noises (an example is represented in Figures 3-2a and 3-2b).

3) **Impulses:** If $\forall f$, $c_{g_r}(f,\varphi,\tau) = \delta_{2F}(\varphi)\delta(\tau)$ and $p_r(t,\varphi,\tau)$ is defined as in Example 1 in Appendix B with $\varepsilon \to 0$, then $y_r(t)$ is a linear combination of impulses (an example is represented in Figures 3-3a and 3-3b).

4) **Damped sinusoids:** If $p_r(t,\varphi,\tau)$ is defined as in Example 1 in Appendix B with $P = 1$ and $\varepsilon \to 0$, and if $c_r(f,\varphi,\tau)$ is defined as in Appendix A with $N_0 = 0$, then $y_r(n)$ follows a real-valued exponential sinusoidal model (ESM) commonly used in HR spectral analysis of time series [8] (an example is represented in Figures 3-4a and 3-4b).

The PTFSFD model also includes some NMF frameworks. Indeed, if $\boldsymbol{W}$ and $\boldsymbol{H}$ are $F \times R$ and $R \times T$ nonnegative matrices of entries W_{fr} and H_{rt}, if $w_r(f,t)$ is 2D-white *Gaussian* noise, and if $p_{\sigma_r}(t,\varphi,\tau) = \sqrt{H_{rt}}\delta_{2F}(\varphi)\delta(\tau)$ then:

5) if $c_{g_r}(f,\varphi,\tau)$=$\sqrt{W_{fr}}\delta_{2F}(\varphi)\delta(\tau)$, $y(f,t)$ follows an **IS-NMF** model [4] of order R, defined by the NMF $\boldsymbol{V}$=$\boldsymbol{W}\boldsymbol{H}$;

6) if $c_r(f,\varphi,\tau) = \sqrt{W_{fr}}\delta_{2F}(\varphi)a_{fr}(\tau)$ where $a_{fr}(\tau)$ is the impulse response of a causal and stable autoregressive filter, $y(f,t)$ follows an **HR-NMF** model [9, 10] of order R, defined by the NMF $\boldsymbol{V}$=$\boldsymbol{W}\boldsymbol{H}$ and the autoregressive filters a_{fr}.

Unlike cases (1)-(4), IS-NMF and HR-NMF are *inconsistent* in the sense defined in Remarks 1 and 2 in Sections 3 and 4. Besides, IS-NMF cannot accurately generate damped sinusoids as in Figure 3-4a, and neither IS-NMF nor HR-NMF can accurately generate impulses as in Figure 3-3a.

7. CONCLUSIONS

In this paper, we proposed a new probabilistic model of non-stationary signals, which consists of a sum of source-filter models. Each component of the sum is obtained by successively applying a multiplication and a convolution to white noise. In order to properly express this model in the TF domain, we used a critically sampled cosine-modulated filter bank, which allows an accurate implementation in the TF domain of convolutions and multiplications in the original time domain, by means of 2D-filters. We then proposed an ARMA parametrisation of these 2D-filters, which can represent a variety of transformations, ranging from smooth to sharply selective in time or in frequency. We have demonstrated that the resulting probabilistic time-frequency source-filter decomposition (PTFSFD) model can represent a broad range of stationary and non-stationary signals, including ARMA processes and uncorrelated noise, and it has both high spectral

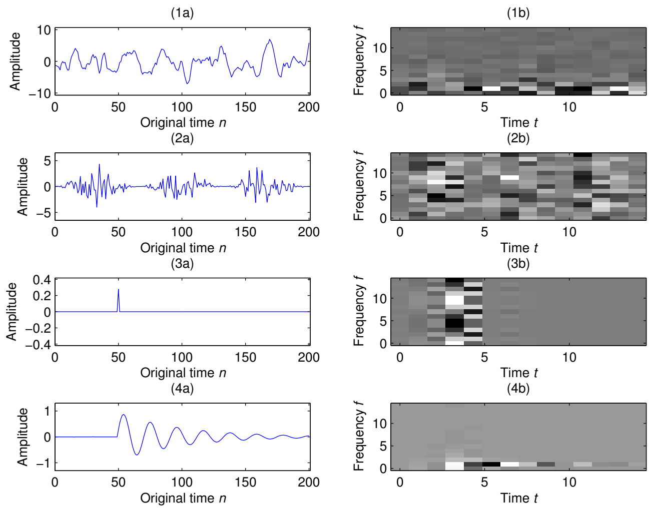

Fig. 3. Examples of signals generated by the PTFSFD model (F=15, L=8, K=4; $y_r(n)$ are in the left column and $y_r(f,t)$ in the right column): (1) is an ARMA process, (2) is an uncorrelated noise, (3) is an impulse, (4) is a damped sinusoid[3]

and temporal resolutions. PTFSFD also generalizes and improves both IS-NMF and HR-NMF: it can accurately generate damped sinusoids, which IS-NMF cannot; and importantly it can also accurately generate impulses, which is not possible for either IS-NMF or HR-NMF.

Because audio signals are sparse in the time-frequency domain, we observed that the PTFSFD model involves a small number of non-zero parameters in practice. In future work, we will investigate enforcing this property, for instance by introducing an a priori distribution of the parameters inducing sparsity. The proposed approach could also be applied to other types of filter banks, *e.g.* STFT filter banks. We also plan to propose some algorithms for estimating this model from the samples of an input signal. In particular, the variational Bayesian methods already investigated in [11] for the HR-NMF model seem to be a promising approach, because of their low complexity and their capacity to efficiently estimate complex graphical models.

Acknowledgements. This work was undertaken while Roland Badeau was visiting the Centre for Digital Music, partly funded by EPSRC Platform Grant EP/K009559/1, and by the French ANR as a part of the DReaM project (ANR-09-CORD-006-03). Mark D. Plumbley is funded by EPSRC Leadership Fellowship EP/G007144/1.

A. ARMA MODEL OF CONVOLUTION

We consider the TF implementation of convolution presented in Section 3, and we define $g(n)$ as the impulse

[3]The numerical values of the model parameters that we used to generate those signals are listed in our Matlab code, which is available upon request.

response of a causal and stable ARMA filter, having only simple poles. Then the partial fraction expansion of its transfer function shows that it can be written in the form $g(n) = g_0(n) + \sum_{k=1}^{P} g_k(n)$, where $P \in \mathbb{N}$, $g_0(n)$ is a causal sequence of support $[0 \dots N_0 - 1]$ (with $N_0 \in \mathbb{N}$), and $\forall k \in [1 \dots P]$, $g_k(n) = A_k e^{\delta_k n} \cos(2\pi\nu_k n + \psi_k) 1_{n \geq 0}$, where $A_k > 0$, $\delta_k < 0$, $\nu_k \in [0, \frac{1}{2}]$, $\psi_k \in \mathbb{R}$. Then $\forall f$, equation (2) yields $c_g(f, \varphi, \tau) = \sum_{k=0}^{P} c_{g_k}(f, \varphi, \tau)$ with $c_{g_0}(f, \varphi, \tau) = (h_f * \tilde{h}_{f-\varphi} * g_0)(F(\tau + L))$ and $\forall k \in [1 \dots P]$, $c_{g_k}(f, \varphi, \tau) = e^{\delta_k F \tau}(A_k(f, \varphi, \tau) \cos(2\pi\nu_k F\tau) + B_k(f, \varphi, \tau) \sin(2\pi\nu_k F\tau))$, where we defined $A_k(f, \varphi, \tau) = A_k \sum_{n=-N+1}^{N-1} (h_f * \tilde{h}_{f-\varphi})(n+N) e^{-\delta_k n} \cos(2\pi\nu_k n - \psi_k) 1_{n \leq F\tau}$ and $B_k(f, \varphi, \tau) = A_k \sum_{n=-N+1}^{N-1} (h_f * \tilde{h}_{f-\varphi})(n+N) e^{-\delta_k n} \sin(2\pi\nu_k n - \psi_k) 1_{n \leq F\tau}$. It can be easily proved that:

- the support of $c_{g_0}(f, \varphi, \tau)$ is [-L+1 … L+$\lceil \frac{N_0-2}{F} \rceil$] w.r.t. τ,
- if $\tau \leq -L$, then $c_{g_0}(f, \varphi, \tau)$, $A_k(f, \varphi, \tau)$ and $B_k(f, \varphi, \tau)$ are zero, thus $c_g(f, \varphi, \tau) = 0$,
- if $\tau \geq L$, $A_k(f, \varphi, \tau)$ and $B_k(f, \varphi, \tau)$ do not depend on τ.

It results that $\forall f, \varphi$, $c_g(f, \varphi, \tau - L + 1)$ is the impulse response of a causal and stable ARMA filter.

As a particular case, suppose that $\forall k \in [1 \dots P]$, $|\delta_k| \ll 1$. If $\tau \geq L$, then $A_k(f, \varphi, \tau)$ and $B_k(f, \varphi, \tau)$ can be neglected as soon as ν_k does not lie in the supports of both $H_f(\nu)$ and $H_{f-\varphi}(\nu)$. Thus for each f and φ, there is a limited number $P(f, \varphi) \leq P$ (possibly 0) of $c_{g_k}(f, \varphi, \tau)$ which contribute to $c_g(f, \varphi, \tau)$. In the general case, we can still consider without loss of generality that $\forall f, \varphi$, there is a limited number $P(f, \varphi) \leq P$ of $c_{g_k}(f, \varphi, \tau)$ which contribute to $c_g(f, \varphi, \tau)$. We then define $Q_a \triangleq 2 \max_{f,\varphi} P(f, \varphi)$ and $Q_b = 2L + Q_a - 1 + \lceil \frac{N_0-2}{F} \rceil$. Then it can be easily proved that $c_g(f, \varphi, \tau)$ matches Definition 1.

B. ARMA MODEL OF MULTIPLICATION

We consider the TF implementation of multiplication presented in Section 4. Suppose that $\sigma_0(n)$ is a band-limited sequence, so that there is $0 < \nu_0 < \frac{1}{2} - \frac{K}{2F}$ such that the DTFT $\Sigma_0(\nu)$ is zero $\forall \nu \in [\nu_0, \frac{1}{2}]$ (we note that if $\nu_0 \ll 1$, then $\sigma_0(n)$ has slow temporal variations). Then the DTFT of the sequence $\sigma_0(Ft - m)h(m)h(m - F\tau)$ is zero $\forall \nu \in [\nu_0 + \frac{K}{2F}, \frac{1}{2}]$. Consequently, equation (5) yields $p_{\sigma_0}(t, \varphi, \tau) = 0$ if $K + \lceil 2F\nu_0 \rceil \leq |\varphi| \leq F$. Besides, let $P \in \mathbb{N}$, and $\forall k \in [1 \dots P]$, let $\sigma_k(n) = \frac{\sigma'_k(n)}{2(\cos(\frac{\pi(n-\phi)}{F}) - \cos(\theta_k))}$, where $\frac{F\theta_k}{\pi} \in \mathbb{R} \backslash \mathbb{Z}$ and $\sigma'_k(n)$ is a band-limited sequence, such that the DTFT $\Sigma'_k(\nu)$ is zero $\forall \nu \in [\nu_0, \frac{1}{2}]$. Then equation (5) yields $p_{\sigma_k}(t, \varphi + 1, \tau) - 2\cos(\theta_k) p_{\sigma_k}(t, \varphi, \tau) + p_{\sigma_k}(t, \varphi - 1, \tau) = p_{\sigma'_k}(t, \varphi, \tau)$. Finally, we define $\sigma(n) = \sum_{k=0}^{P} \sigma_k(n)$.

Example 1 (Linear combination of impulses). As an example, suppose that $\sigma_0(n) = 0$, and $\forall k \in [1 \dots P]$, $\theta_k = \frac{\pi(\phi - n_k)}{F} + \varepsilon$ (where $n_k \in \mathbb{Z}$, all n_k are distinct, and $\varepsilon > 0$), and $\sigma'_k(n) = 2A_k \varepsilon \sin(\frac{\pi(2\phi - n - n_k)}{2F} + \frac{\varepsilon}{2}) \text{sinc}(\frac{n - n_k}{2F})$ (where $\text{sinc}(u) = \frac{\sin(\pi u)}{\pi u}$). In particular, $\nu_0 = \frac{3}{4F}$. Then when $\varepsilon \to 0$, $\sigma_k(n) \to A_k \delta(n - n_k)$. Asymptotically, we thus get $\sigma(n) = \sum_{k=1}^{P} A_k \delta(n - n_k)$, *i.e.* a linear combination of impulses. Moreover, in this case, if either t or $t - \tau \notin \left[\lceil \frac{n_k}{F} \rceil \dots \lceil \frac{n_k}{F} \rceil + L - 1\right]$, then $p_{\sigma_k}(t, \varphi, \tau) = 0$. Thus for each t and τ, there is a finite number $P(t, \tau) \leq P$ (possibly 0) of $p_{\sigma_k}(t, \varphi, \tau)$ which contribute to $p_{\sigma}(t, \varphi, \tau)$.

In the general case, we can still consider without loss of generality that $\forall t, \tau$, there is a limited number $P(t, \tau) \leq P$ of $p_{\sigma_k}(t, \varphi, \tau)$ which contribute to $p_{\sigma}(t, \varphi, \tau)$. Moreover, we suppose that $P_\alpha = \max_{t,\tau} P(t, \tau)$ and ν_0 are such that $P_\alpha < F - K$ and $\nu_0 < \frac{1}{2} - \frac{K + P_\alpha}{2F}$, and we define $P_\beta = K + P_\alpha - 1 + \lceil 2F\nu_0 \rceil < F$. Then it can be easily proved that $p_{\sigma}(t, \varphi, \tau)$ matches Definition 2.

C. REFERENCES

[1] M. N. Schmidt and H. Laurberg, "Non-negative matrix factorization with Gaussian process priors," *Computational Intelligence and Neuroscience*, vol. 2008, 2008, Article ID 361705, 10 pages.

[2] P. Smaragdis, "Probabilistic decompositions of spectra for sound separation," in *Blind Speech Separation*, S. Makino, T.-W. Lee, and H. Sawada, Eds., pp. 365–386. Springer, 2007.

[3] T. Virtanen, A. Cemgil, and S. Godsill, "Bayesian extensions to non-negative matrix factorisation for audio signal modelling," in *Proc. ICASSP*, Las Vegas, Nevada, USA, Apr. 2008, pp. 1825–1828, IEEE.

[4] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[5] J. Le Roux, H. Kameoka, N. Ono, and S. Sagayama, "Fast signal reconstruction from magnitude STFT spectrogram based on spectrogram consistency," in *Proc. DAFx*, Graz, Austria, Sept. 2010, pp. 397–403.

[6] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 217–220, Mar. 2013.

[7] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[8] M. H. Hayes, *Statistical Digital Signal Processing And Modeling*, Wiley, Aug. 2009.

[9] R. Badeau, "Gaussian modeling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF)," in *Proc. WASPAA*, New York, USA, Oct. 2011, pp. 253–256.

[10] R. Badeau, "High resolution NMF for modeling mixtures of non-stationary signals in the time-frequency domain," Tech. Rep. 2012D004, Telecom ParisTech, Paris, France, July 2012.

[11] R. Badeau and A. Drémeau, "Variational Bayesian EM algorithm for modeling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)," in *Proc. ICASSP*, Vancouver, Canada, May 2013, pp. 6171–6175.