Non-stationary audio signals analysis for the joint estimation of deformations and spectrum

Adrien Meynard

Aix Marseille Univ, CNRS, Centrale Marseille, I2M, UMR 7373, Marseille, France.

I. Introduction

Abstract

An approach to the spectral estimation for some classes of non-stationary random signals is developed, that addresses stationary random processes deformed by a stationarity-breaking transformation. Examples include time warping, amplitude and frequency modulations, and others. Under suitable smoothness assumptions on the transformation, approximate expressions are obtained in adapted representation spaces. In the Gaussian case, this leads to approximate maximum likelihood estimation algorithms, which are illustrated on synthetic as well as real signals.

II. Model and approximations

Model

We consider the following breaking-stationarity deformation operators:

- Amplitude modulation: $\mathcal{A}_a x(t) = a(t)x(t)$,
- Time warping: $\mathcal{D}_\gamma x(t) = \sqrt{\gamma'(t)}x(\gamma(t))$.

Assume X is a zero mean, Gaussian stationary generalized random process with spectrum denoted by $\mathcal{S}_X$. The observation is of the form

$$Y = \mathcal{A}_a \mathcal{D}_\gamma X \ .$$

The wavelet representation allows highlighting the effect of the deformation on the stationary signal. It is defined as

$$\mathcal{W}_x(s,\tau) = \langle x, \psi_{s\tau}\rangle, \text{ where } \psi_{s\tau}(t) = q^{-s/2}\psi\left(q^{-s}(t-\tau)\right) \ ,$$

where ψ is an analytic wavelet (i.e. $\hat{\psi}(\omega) = 0, \forall\omega < 0$).

Approximation theorem

$$\widetilde{\mathcal{W}}_Y(s,\tau) \approx \mathcal{W}_Y(s,\tau) = a(\tau)\mathcal{W}_X\left(s + \log_q(\gamma'(\tau)), \gamma(\tau)\right)$$

The error $\varepsilon = \mathcal{W}_Y - \widetilde{\mathcal{W}}_Y$ is a zero mean Gaussian random field whose variance $\mathbb{E}\left\{|\varepsilon(s,\tau)|^2\right\}$ depends on the smoothness of a and γ', and on the wavelet decay rate.

III. Estimation procedure and algorithm

Joint spectrum and deformation estimation algorithm

The joint estimation of the deformation operator and the spectrum of the underlying stationary signal rests on two steps:

- Estimation of the deformation assuming the spectrum is known,
- Estimation of the spectrum assuming the deformation is known.

$\Rightarrow$ These two steps are computed alternatively until convergence of the estimators.

a) Step 1: Deformation estimation.

Assume that the spectrum of the underlying stationary signal $\mathcal{S}_X$ is known.

At fixed time τ, the parameters Θ to estimate are

$$\Theta_n = (\theta_{n,1}, \theta_{n,2}) := (a(\tau_n)^2, \log_q(\gamma'(\tau_n))).$$

Denote by $\mathbf{w}_{y,\tau_n}$ the restriction of $\widetilde{\mathcal{W}}_y(\cdot,\tau_n)$ to a finite sampling subset of the scale space. $\mathbf{w}_{y,\tau_n}$ is a zero mean circular, Gaussian random vector. This yields the following log-likelihood

$$\mathcal{L}(\Theta_n) = -\ln|\det \mathbf{C}(\Theta_n)| - \mathbf{C}(\Theta_n)^{-1}\mathbf{w}_{y,\tau_n}\cdot\mathbf{w}_{y,\tau_n} \ ,$$

where

$$\mathbf{C}(\Theta_n)_{ij} = \theta_{n,1}q^{(s_i+s_j)/2}\int_0^\infty \mathcal{S}_X(q^{-\theta_{n,2}}\xi)\overline{\hat{\psi}}(q^{s_i}\xi)\hat{\psi}(q^{s_j}\xi)d\xi \ .$$

Then, the log-likelihood is maximized with respect to each component of Θ_n.

b) Step 2: Spectrum estimation.

Assume that the deformation operators $\theta_{n,1}$ and $\theta_{n,2}$ are known for all n.

Construct the rectified wavelet transform $\mathbf{w}_{x,s_k}[n] = \theta_{n,1}{}^{-1/2}\widetilde{\mathcal{W}}_y(s_k - \theta_{n,2}, \tau_n)$, then:

$$\mathbb{E}\left\{\frac{1}{N_\tau\|\psi\|_2^2}\|\mathbf{w}_{x,s_k}\|^2\right\} = \frac{1}{\|\psi\|_2^2}\int_0^\infty \mathcal{S}_X(\xi)q^{s_k}\left|\hat{\psi}\left(q^{s_k}\xi\right)\right|^2 d\xi \ ,$$

which is a filtered version of $\mathcal{S}_X$ around frequency $\omega_k = q^{-s_k}\omega_0$ where ω_0 is the central frequency of $|\hat{\psi}|^2$. The considered estimator is obtained by replacing the expectation by the sample variance of the vector $\mathbf{w}_{x,s_k}$.

Performances: Cramér-Rao lower bound and Slepian-Bangs formula

For any unbiased estimator $\hat{\theta}_{n,i}$ ($i = 1$ or 2) of a component $\theta_{n,i}$ of the bivariate parameter Θ_n,

$$\mathbb{E}\left\{(\hat{\theta}_{n,i} - \theta_{n,i})^2\right\} \geq \text{CRLB}(\theta_{n,i}).$$

When the observation is zero mean complex Gaussian

$$\text{CRLB}(\theta_{n,i}) = \left(\text{Trace}\left\{\left(\mathbf{C}(\Theta_n)^{-1}\frac{\partial\mathbf{C}(\Theta_n)}{\partial\theta_{n,i}}\right)^2\right\}\right)^{-1} \ .$$

Remark: One can evaluate the robustness to noise performances of the estimators and show that the $\theta_{n,2}$ estimation is robust to additive Gaussian white noise while the estimation of $\theta_{n,1}$ is rather sensitive to additive noise.

In the following numerical examples, the sharp wavelet $\psi_\sharp$ (with infinitely many vanishing moments) is used, defined in the positive Fourier domain by

$$\hat{\psi}_\sharp(\omega) = \epsilon^{\frac{\delta(\omega,\omega_0)}{\delta(\omega_1,\omega_0)}} \ , \ \omega > 0 \text{ where } \delta(a,b) = \frac{1}{2}\left(\frac{a}{b} + \frac{b}{a}\right) - 1 \ .$$

Here ω_0 is the mode of $\hat{\psi}$, ω_1 is chosen so that $\hat{\psi}_\sharp(\omega_1) = \epsilon$.

Some numerical results

Performances of the algorithm are evaluated on a synthetic toy signal comparing the results with the true values of the estimated parameters.

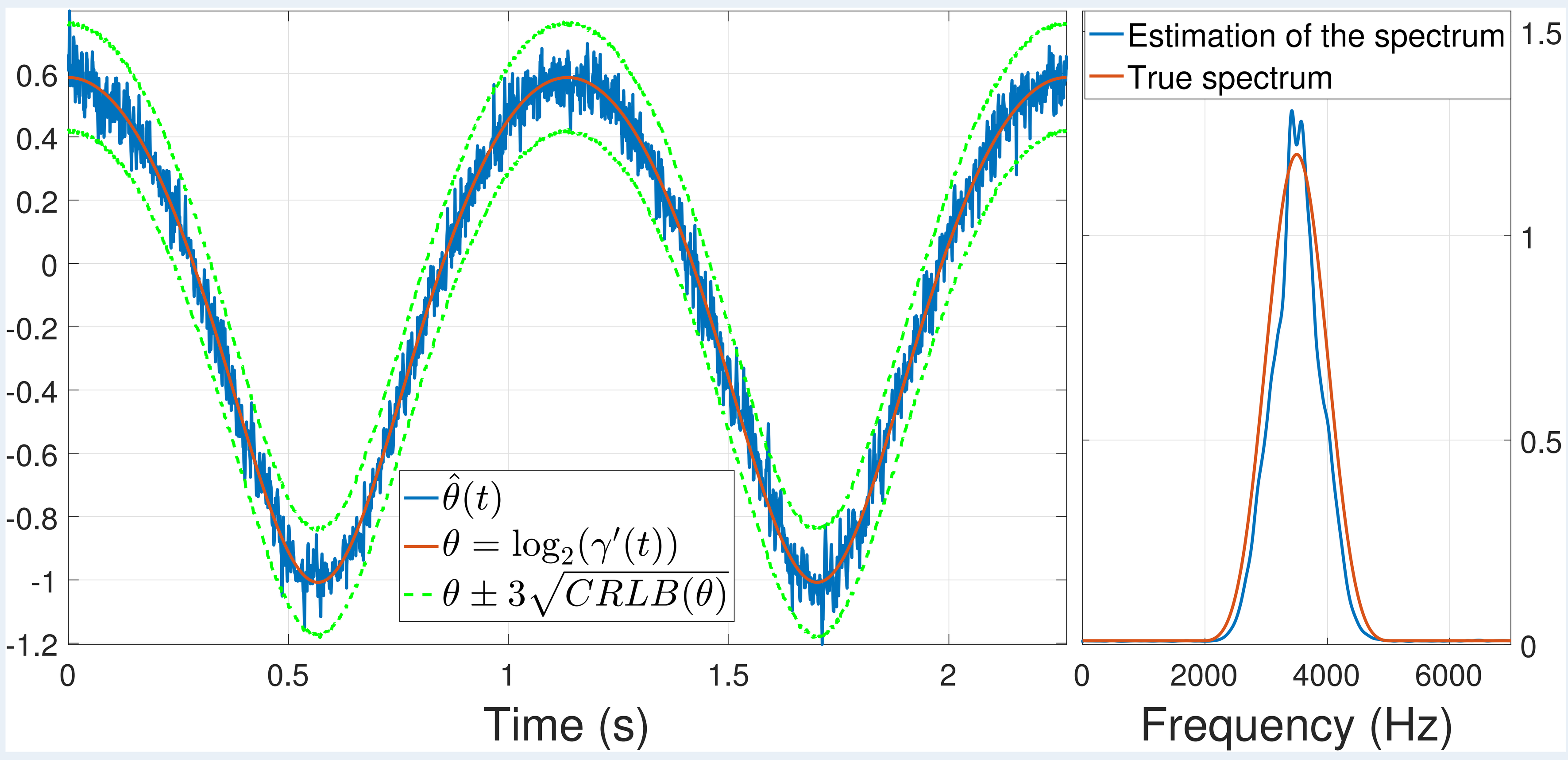

Figure: Joint time warping/spectrum estimation on a synthetic signal. Left: time warping function estimate (full, blue), ground truth (full, red) and Cramér-Rao bound (dotted, green); Right: spectrum of the underlying stationary signal.

The algorithm has been tested on real life audio signals. We display here the example of a singing female voice signal.

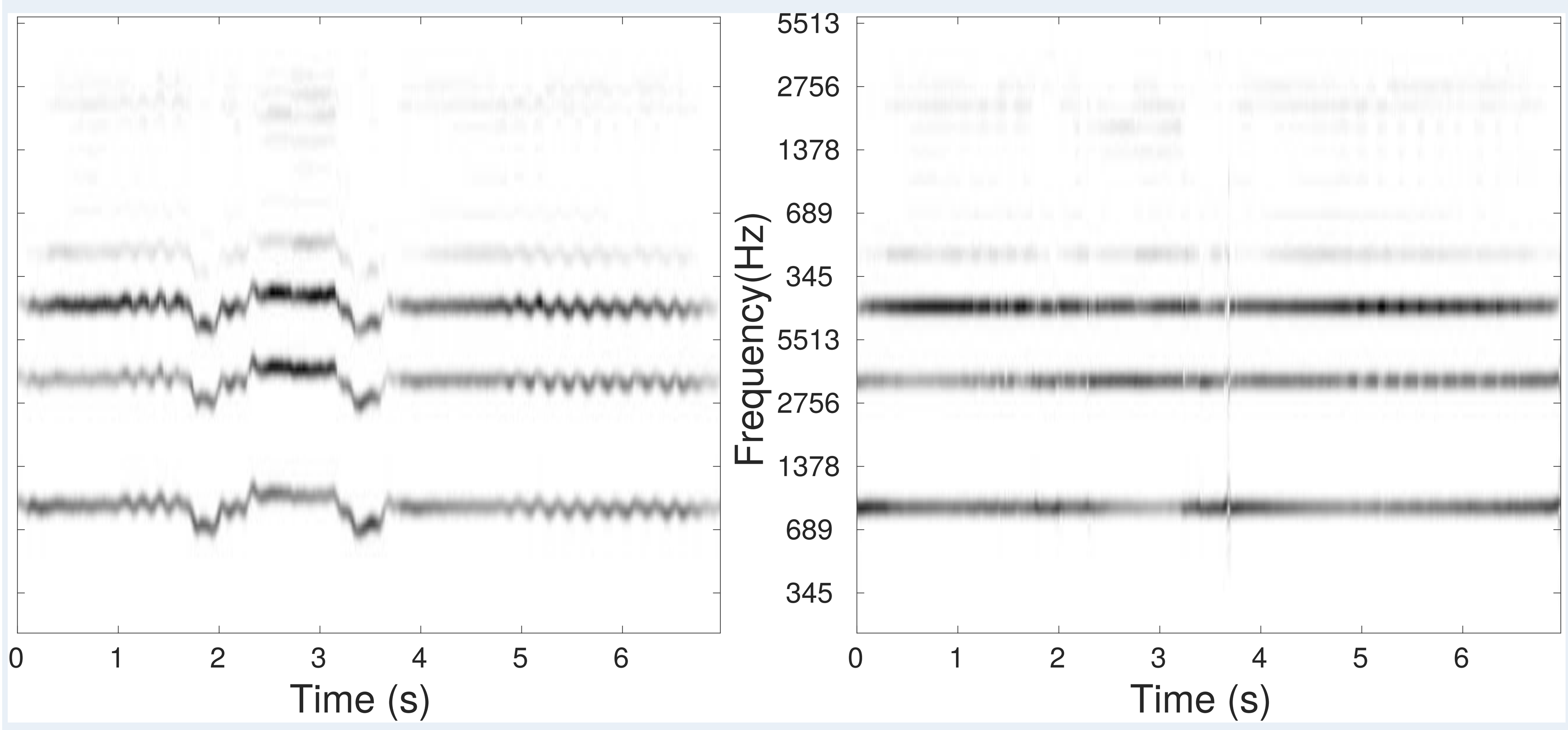

Figure: Scalograms of the original signal and the estimated underlying stationary signal.

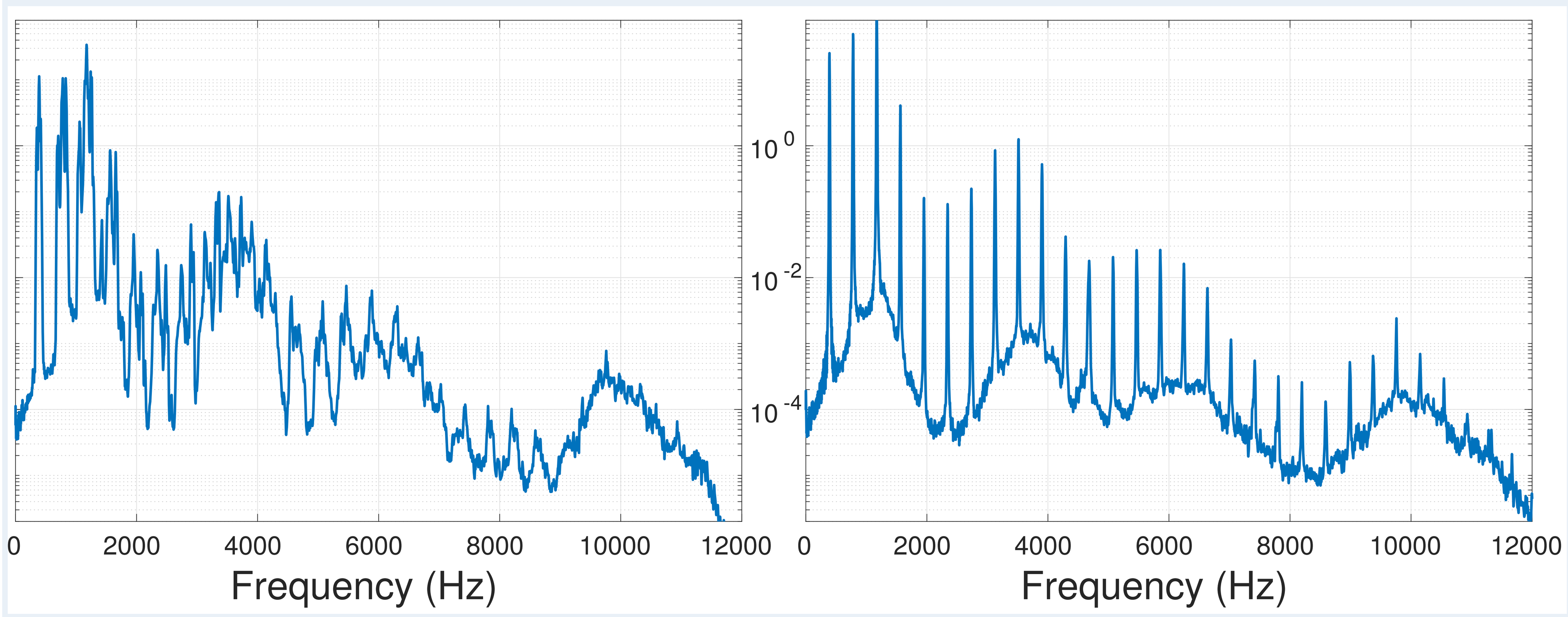

Figure: Estimation of the power spectra of the input signal (left) and the processed signals (right).

As a stationarity check, the harmonic structure of the estimated spectrum is remarkably neat.

IV. Conclusion and perspectives

Applications

- Analysis of non stationary signals to extract spectral properties.
- Synthesis of new signals applying any deformation to "stationarized" signals.

Perspectives and further work

- Joint estimation of frequency modulation and time warping using an adapted transform (see [1]).
- Estimation of deformations and spectrum for signals where the Gaussian assumption on the underlying stationary signal does not hold.

References

Adrien Meynard and Bruno Torrésani.
Spectral estimation for non-stationary signal classes.
In *Sampling Theory and Applications*, Proceedings of SampTA17, Tallinn, Estonia, July 2017.

Maureen Clerc and Stéphane Mallat.
Estimating deformations of stationary processes.
Ann. Statist., 31(6):1772–1821, 12 2003.

Harold Omer and Bruno Torrésani.
Time-frequency and time-scale analysis of deformed stationary processes, with application to non-stationary sound modeling.
Applied and Computational Harmonic Analysis, 43(1):1 – 22, 2017.

Adrien Meynard and Bruno Torrésani.
Spectral analysis for non-stationary audio.
In preparation.