

Consistent Spectrogram Separation from Nonstationary Mixture

Adrien Meynard

ENS de Lyon, France

adrien.meynard@ens-lyon.fr

Marina Krémé

Mines Saint-Etienne, France

marina.kreme@emse.fr



1. Motivation and Context

Why spectrogram separation?

Many signal-processing tasks require spectrogram decomposition:

- Audio source separation
- Sound event detection
- Speech enhancement
- Bioacoustic signal analysis

◆ Most approaches focus on reconstructing the source signals.

◆ **Our objective:**

→ **Direct spectrogram separation from a single mixture.**

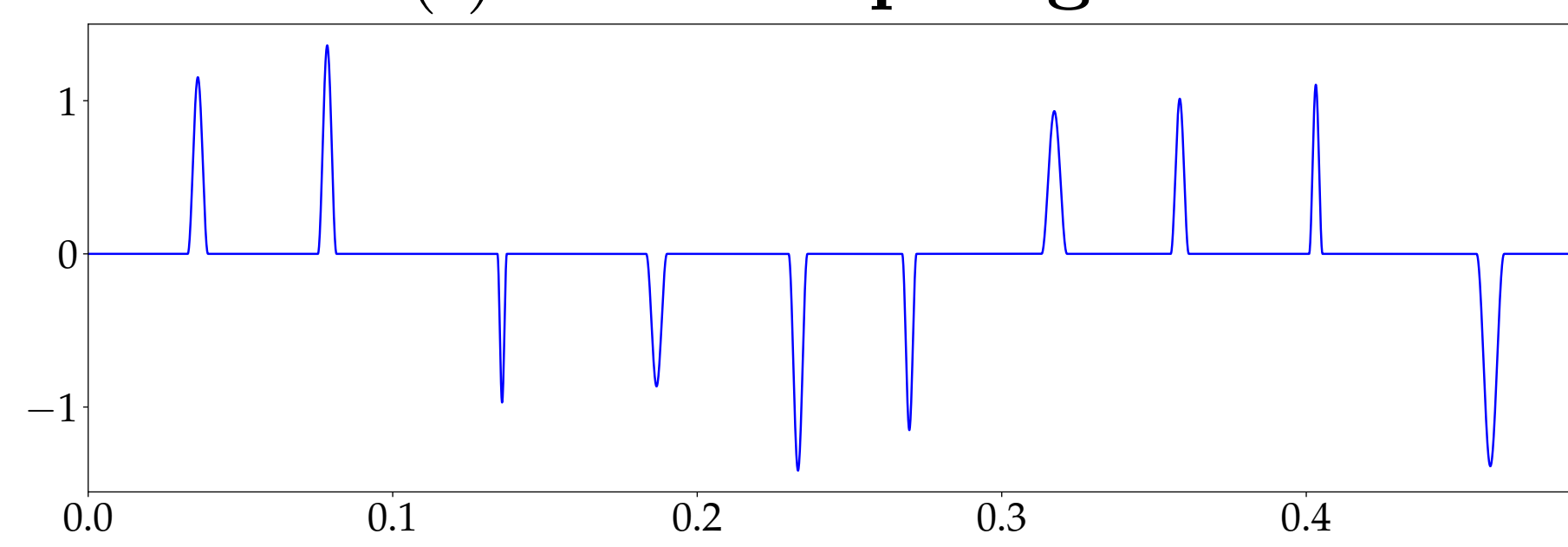
◆ **Strategy:** Inspired by image-processing techniques [1].

2. Signal Model

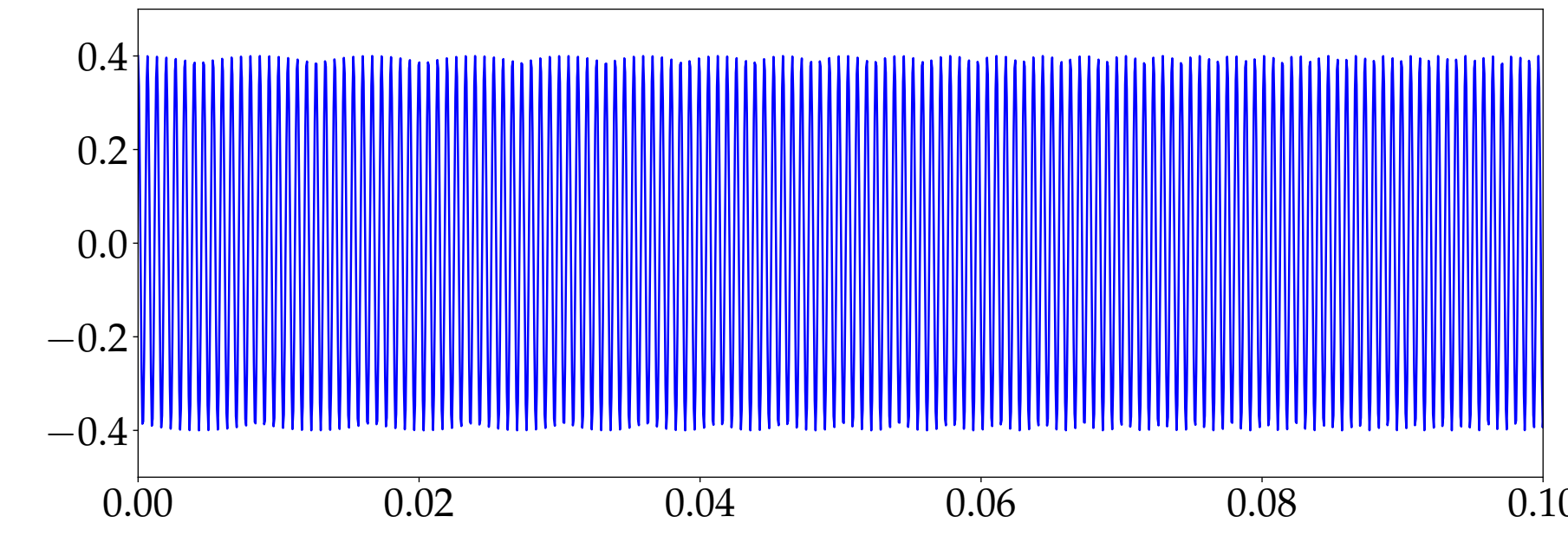
Mixture

$$z(t) = x(t) + y(t)$$

$x(t)$ is a bumps signal.



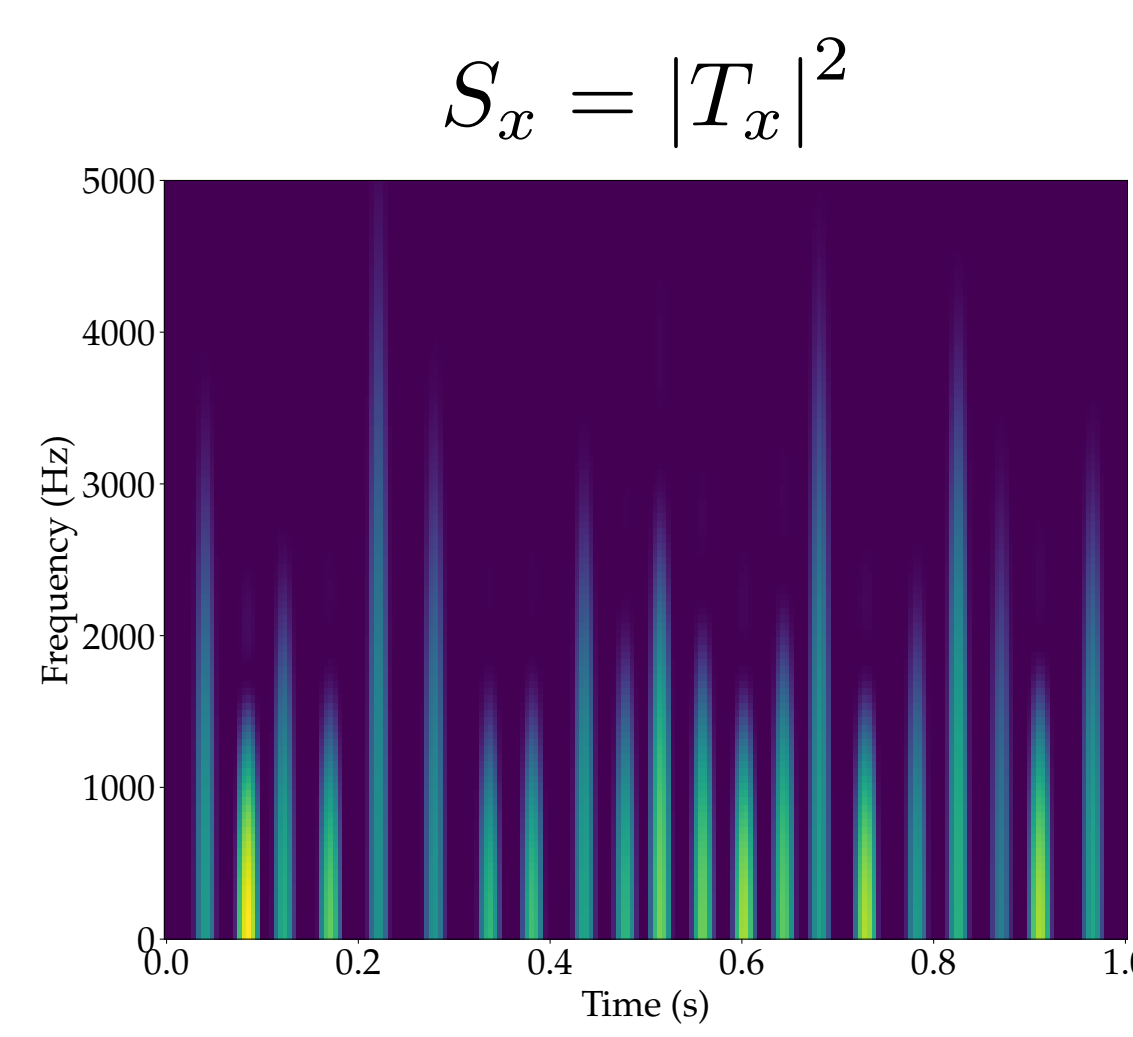
$y(t)$ is a multicomponent AM-FM signal.



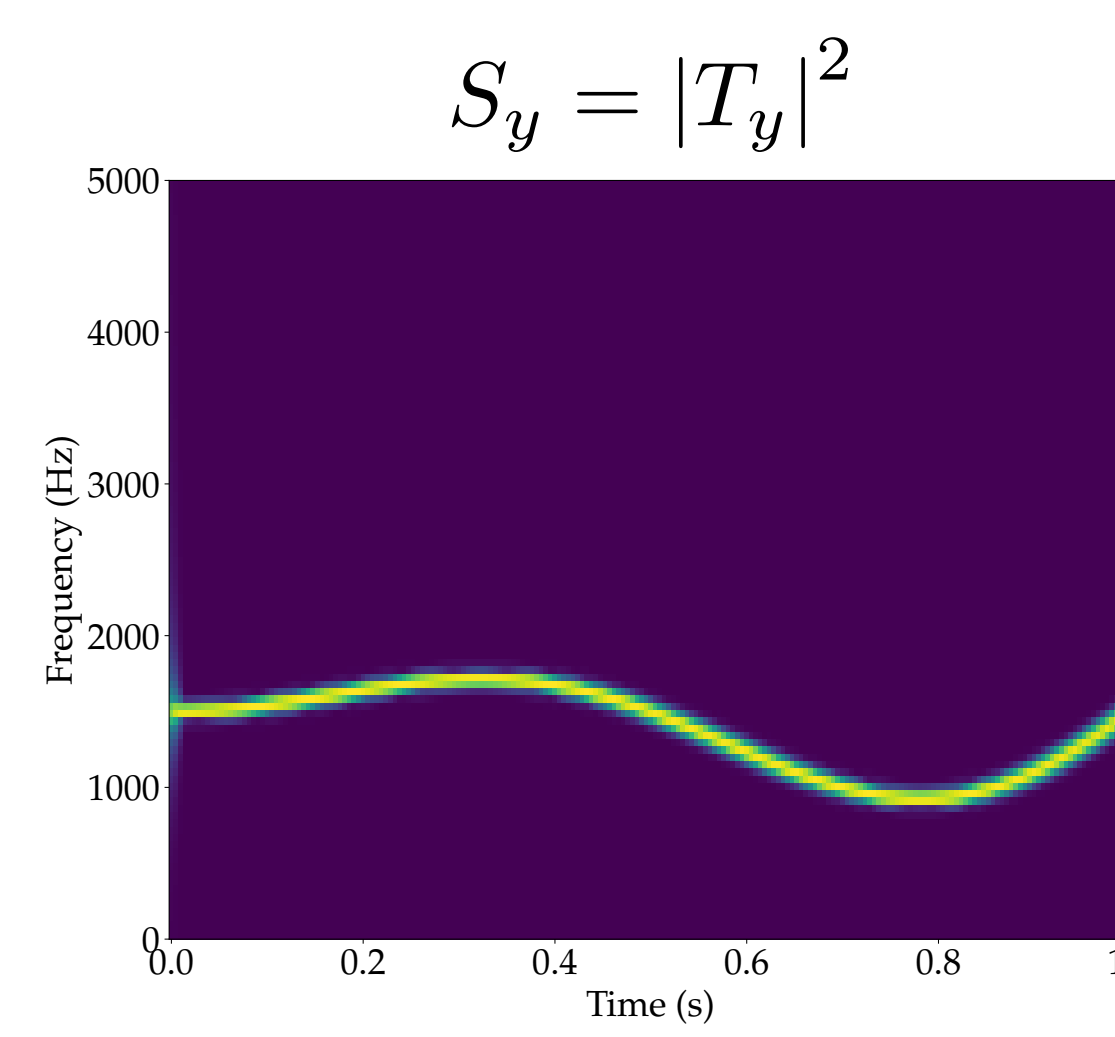
3. Time-Frequency Analysis

Short-Time Fourier Transform (STFT)

$$T_x(\nu, \tau) = \int_{\mathbb{R}} x(t)g(t - \tau)e^{-i2\pi\nu t}dt.$$

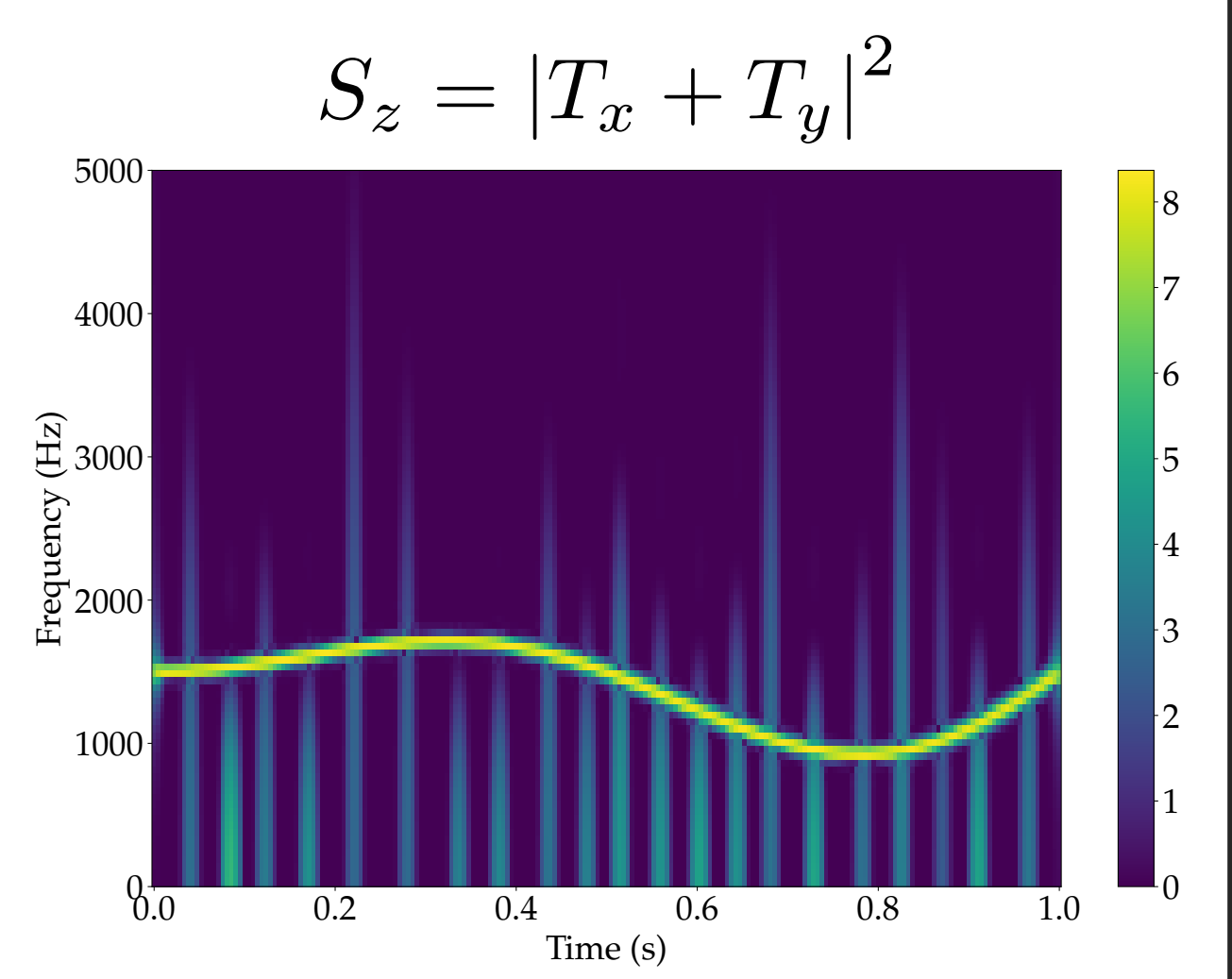


S_x varies slowly along the frequency axis.



Asymptotic properties

S_y is precisely localized around the instantaneous frequencies [2].



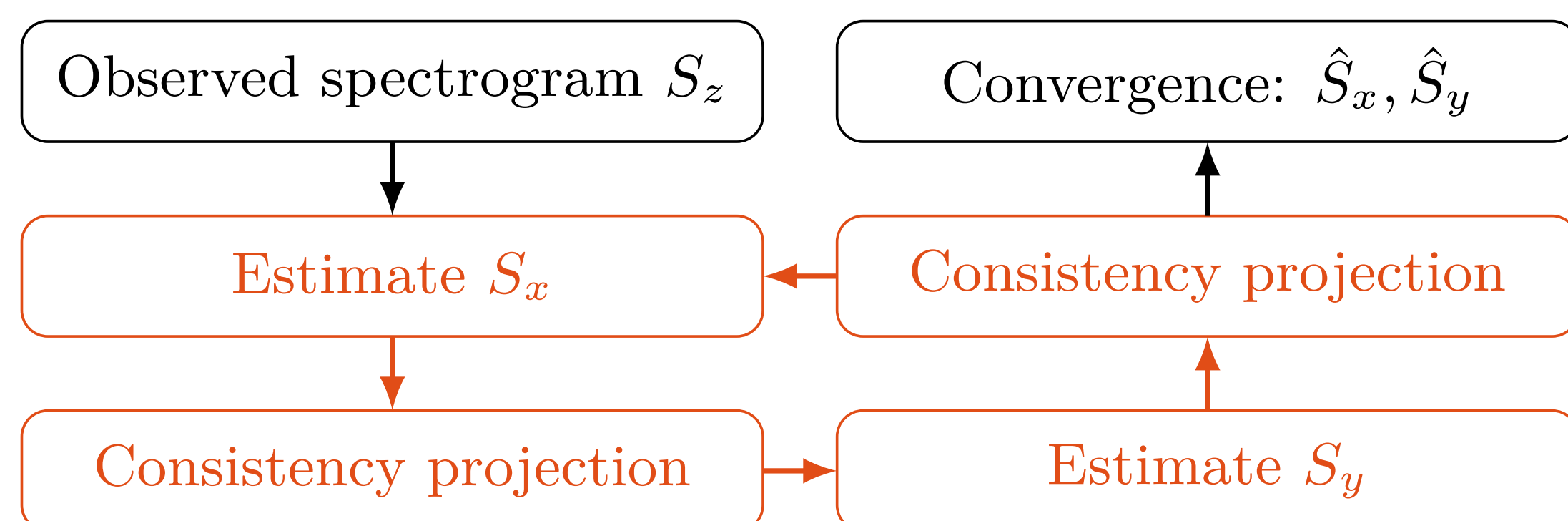
$S_z = S_x + S_y + \text{interference}$

4. Proposed Method

A. Inverse Problem Formulation:

$$\hat{S}_x, \hat{S}_y = \arg \min_{S_x, S_y} \|S_z - (S_x + S_y)\|_2^2 + \lambda \|\partial_\nu S_x\|_2^2 + \mu \|S_y\|_1$$

B. Alternating Optimization Algorithm Ensuring Consistency:



◆ **Estimation of S_x :** closed-form solution.

◆ **Estimation of S_y via FISTA algorithm [3].**

Consistency

- ◆ Complex valued-matrix X .
- ◆ Consistency operator: $\Pi = \text{STFT} \circ \text{STFT}^{-1}$.
- ◆ X is *consistent* $\Leftrightarrow \Pi(X) = X$.

Consistency projection: After each update of S_x and S_y .

◆ Phase incorporation:

$$T_x^{(k)} \leftarrow \sqrt{S_x^{(k)}} \arg(T_z) \quad \text{and} \quad T_y^{(k)} \leftarrow \sqrt{S_y^{(k)}} \arg(T_z).$$

◆ Consistency:

$$S_x^{(k)} \leftarrow |\Pi(T_x^{(k)})|^2 \quad \text{and} \quad S_y^{(k)} \leftarrow |\Pi(T_y^{(k)})|^2.$$

5. Results

A. Synthetic Mixture:

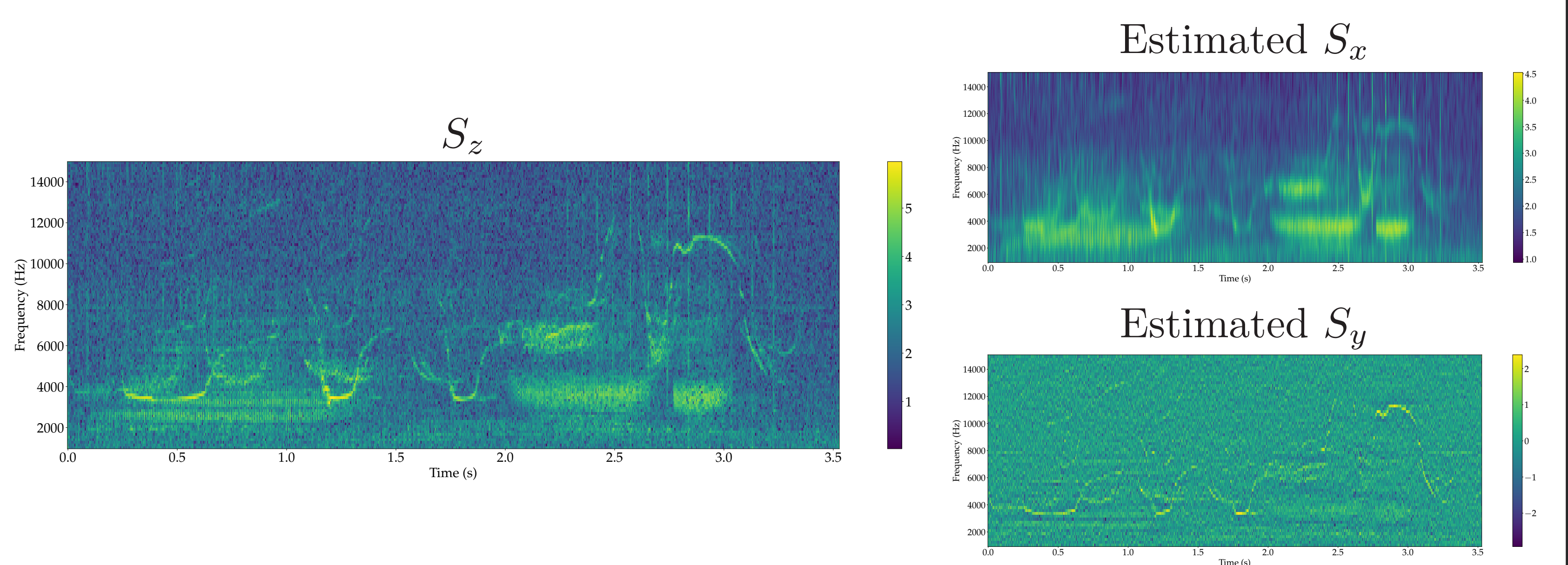
→ 1 sec. synthetic signal with random bumps (arrival time, height, width), and a chirp-like AM-FM component.

Residual norm after convergence	
Proposed Method	NMF
7.8×10^{-5}	1.4×10^{-2}

B. Application to Bioacoustics:

→ 3.5 sec. dolphin recording containing clicks and whistles.

→ $\lambda = 0.1$ and $\mu = 0.00001$: Convergence in 165 iterations.



6. Conclusions and Future Work

Contributions:

- Inverse problem formulation;
- Asymptotic guarantees;
- Integration of the consistency constraint.

Ongoing work:

- Controlling cross-terms;
- Phase reconstruction;
- Signal separation.

7. References

- [1] Aujol, Gilboa, Chan, Osher (2006). Structure-texture image decomposition: modeling, algorithms, and parameter selection.
- [2] Kowalski, Meynard, Wu (2018). Convex optimization approach to signals with fast varying instantaneous frequency.
- [3] Beck, Teboulle (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems.
- [4] Yatabe (2020). Consistent ICA: Determined BSS meets spectrogram consistency.

