Alpha-Stable Multichannel Audio Source Separation

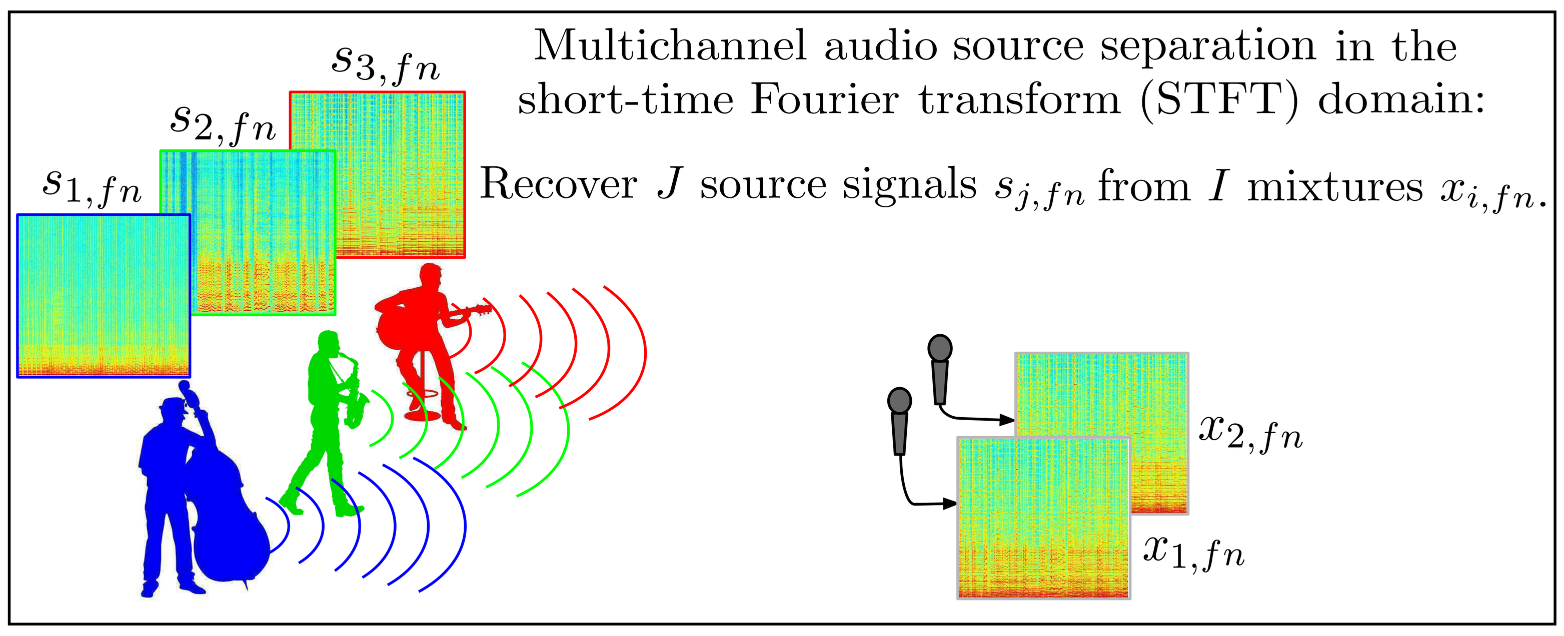
Simon Leglaive[1], Umut Şimşekli[1], Antoine Liutkus[2], Roland Badeau[1], Gaël Richard[1]

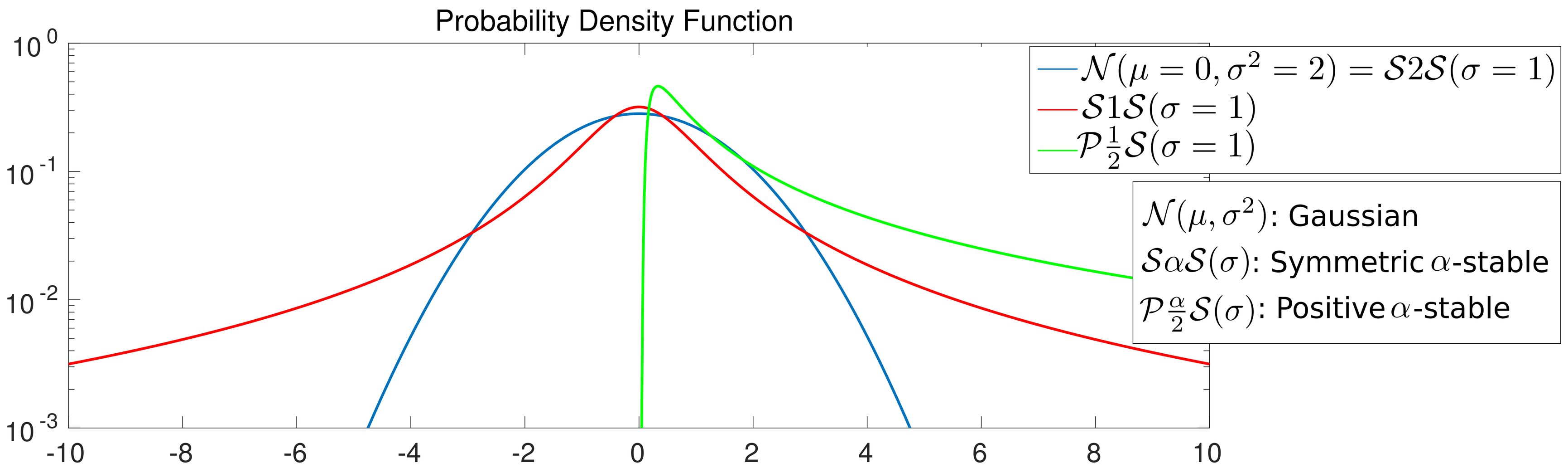
1: LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

2: Inria, Speech Processing Team, Villers-lès-Nancy, France

International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.

1. Introduction





Motivations:

- ▷ Current approaches are based on Gaussian models in which data are constrained to lie just within a few standard deviations from the mean.
- ▷ The proposed approach relies on a heavy-tailed model. It allows us to account for spurious data or important uncertainty in the model.

Contributions:

- ▷ Generalization of the Gaussian source separation framework [1] using the complex multivariate elliptically contoured stable distribution.
- ▷ Inference with the Monte Carlo Expectation Maximization algorithm.
- ▷ Performance improvements in audio source separation under corrupted mixtures and coding-based informed source separation.

2. Model

Gaussian model [1]

▷ Observed mixtures: $\mathbf{x}_{fn} = [x_{1,fn}, ..., x_{I,fn}]^T \in \mathbb{C}^I$;

▷ Latent sources: $\mathbf{s}_{fn} = [s_{1,fn}, ..., s_{J,fn}]^T \in \mathbb{C}^J$.

▷ Joint distribution of the latent and observed variables:

$$\begin{pmatrix}\mathbf{x}_{fn}\\ \mathbf{s}_{fn}\end{pmatrix} \sim \mathcal{N}_c\left(0, \boldsymbol{\Sigma}_{fn} = \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{x},fn} & \mathbf{A}_f\boldsymbol{\Sigma}_{\mathbf{s},fn}\\ \boldsymbol{\Sigma}_{\mathbf{s},fn}\mathbf{A}_f^\star & \boldsymbol{\Sigma}_{\mathbf{s},fn}\end{pmatrix}\right),$$

where $\boldsymbol{\Sigma}_{\mathbf{x},fn} = \mathbf{A}_f\boldsymbol{\Sigma}_{\mathbf{s},fn}\mathbf{A}_f^\star + \boldsymbol{\Sigma}_{\mathbf{b},f}$.

- $\mathbf{A}_f \in \mathbb{C}^{I\times J}$ is the mixing matrix;
- $\cdot^\star$ denotes the Hermitian conjugate;
- $\boldsymbol{\Sigma}_{\mathbf{s},fn}$ is the source covariance matrix parametrized by a non-negative matrix factorization (NMF) model:

$$\boldsymbol{\Sigma}_{s,fn} = \text{diag}([v_{j,fn}]_j), \text{ with } v_{j,fn} = [\mathbf{W}_j\mathbf{H}_j]_{fn} \in \mathbb{R}_+;$$

- $\boldsymbol{\Sigma}_{\mathbf{b},f}$ is the noise covariance matrix.

Proposed elliptically contoured stable model

$$\begin{pmatrix}\mathbf{x}_{fn}\\ \mathbf{s}_{fn}\end{pmatrix} \sim \mathcal{E}\alpha\mathcal{S}_c(\boldsymbol{\Sigma}_{fn}) \iff \begin{cases}\phi_{fn} \sim \mathcal{P}\frac{\alpha}{2}\mathcal{S}\left(2\left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}\right);\\ \begin{pmatrix}\mathbf{x}_{fn}\\ \mathbf{s}_{fn}\end{pmatrix}|\phi_{fn} \sim \mathcal{N}_c(0, \phi_{fn}\boldsymbol{\Sigma}_{fn}),\end{cases}$$

with $\alpha \in (0,2]$ and the same structure of $\boldsymbol{\Sigma}_{fn}$ as in the Gaussian case.

3. Inference

Model parameter set: $\boldsymbol{\Theta} = \{\{\mathbf{W}_j, \mathbf{H}_j\}_j, \{\boldsymbol{\Sigma}_{\mathbf{b},f}, \mathbf{A}_f\}_f\}$.

Monte Carlo Expectation Maximization algorithm

▷ $\mathbf{X} = \{\mathbf{x}_{fn}\}_{f,n}$, $\mathbf{S} = \{\mathbf{s}_{fn}\}_{f,n}$, $\boldsymbol{\Phi} = \{\phi_{fn}\}_{f,n}$.

▷ **Iterative algorithm**:

- E-step: $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}_{\text{old}}) = \mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}_{\text{old}}}[\ln p(\mathbf{X},\mathbf{S},\boldsymbol{\Phi}|\boldsymbol{\Theta})]$;
- M-step: $\boldsymbol{\Theta}_{\text{new}} = \arg\max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}_{\text{old}})$.

▷ The E-step requires the computation of $\hat{\phi}_{fn}^{-1} := \mathbb{E}_{\phi_{fn}|\mathbf{x}_{fn};\boldsymbol{\Theta}}[\phi_{fn}^{-1}]$
$\to$ not available in an analytical form.

▷ **Metropolis-Hastings** algorithm for approximating intractable expectations: compute an average of samples drawn from $p(\phi_{fn}|\mathbf{x}_{fn};\boldsymbol{\Theta})$.

Source posterior moments, i.e. according to $p(\mathbf{s}_{fn}|\mathbf{x}_{fn};\boldsymbol{\Theta})$

▷ $\hat{\mathbf{s}}_{fn} := \mathbb{E}_{\mathbf{s}_{fn}|\mathbf{x}_{fn};\boldsymbol{\Theta}}[\mathbf{s}_{fn}] = \mathbf{G}_{\mathbf{s},fn}\mathbf{x}_{fn}$, $\quad \mathbf{G}_{\mathbf{s},fn} = \boldsymbol{\Sigma}_{\mathbf{s},fn}\mathbf{A}_f^\star\boldsymbol{\Sigma}_{\mathbf{x},fn}^{-1}$.

The source estimate is given by $\hat{\mathbf{s}}_{fn}$ (Wiener filtering). It does not depend on ϕ_{fn} and it is exactly the same as in the Gaussian case.

$$\boldsymbol{\Sigma}_{s,fn}^{post} := \mathbb{E}_{\mathbf{s}_{fn}|\mathbf{x}_{fn};\boldsymbol{\Theta}}\left[(\mathbf{s}_{fn}-\hat{\mathbf{s}}_{fn})(\mathbf{s}_{fn}-\hat{\mathbf{s}}_{fn})^\star\right] = \mathbb{E}_{\phi_{fn}|\mathbf{x}_{fn};\boldsymbol{\Theta}}[\phi_{fn}](\boldsymbol{\Sigma}_{\mathbf{s},fn} - \mathbf{G}_{\mathbf{s},fn}\mathbf{A}_f\boldsymbol{\Sigma}_{\mathbf{s},fn}).$$

The posterior mean of ϕ_{fn} does not admit an analytical form. It is approximated using the Metropolis-Hastings algorithm.

4. Source Separation Under Corrupted Mixtures

▷ **Proof of concept experiment**: A few number of randomly selected time-frequency points are set to very high values in the mixture signals before performing separation ($\to$ highly audible noise).

▷ **Claim**: ϕ_{fn} captures the uncertainty about the model at time-frequency point (f,n). This impulse variable should capture the outliers.

▷ **Modified Wiener filtering** for *canceling* the noise:

$$\hat{\mathbf{s}}_{fn}^{\text{modif}} = \mathbf{G}_{\mathbf{s},fn}\hat{\phi}_{fn}^{-1}\mathbf{x}_{fn}.$$

▷ **Dataset**: 8 stereo mixtures. The duration ranges from 12 to 28 seconds. Each mixture contains between 2 and 4 instrumental sources.

▷ **Performance measure**: Signal-to-Distortion Ratio (SDR) in dB.

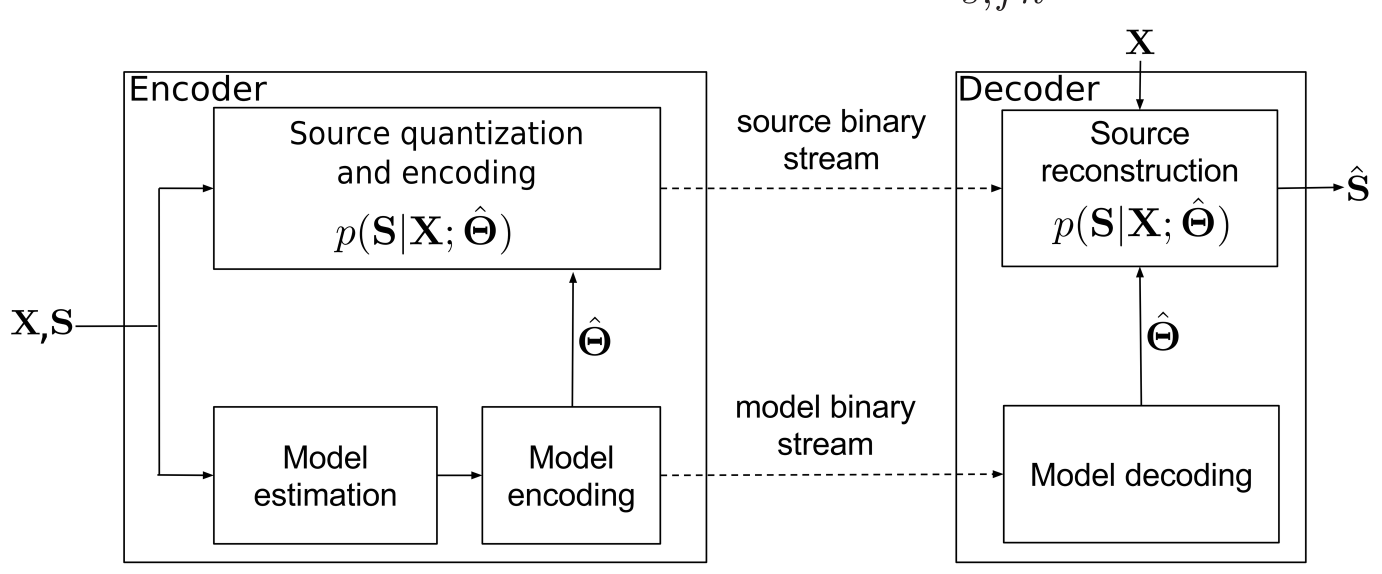
▷ **Results**: Gaussian model with standard Wiener filtering vs. elliptically contoured stable model with the modified estimation procedure ($\alpha = 1.5$):

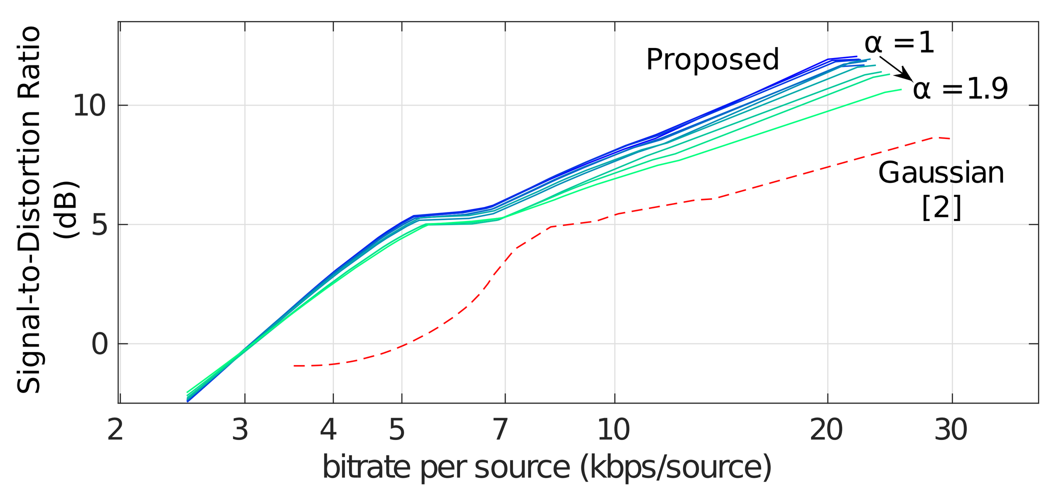
Initialization	Oracle		Blind	
Model	Gaussian	α-stable	Gaussian	α-stable
SDR (dB)	-3.7	**4.5**	-6.7	**0.6**

5. Coding-based Informed Source Separation

▷ **Model stream**: Parameters representing the source posterior means.

▷ **Source stream**: Error between the true source coefficients and their posterior means, encoded through the use of $\boldsymbol{\Sigma}_{s,fn}^{post}$.





6. Conclusion

▷ Using the elliptically contoured stable distribution allows us to take uncertainty about the model into account.

▷ Future work: Justified source estimator that uses the impulse variables.

7. References

[1] A. Ozerov and C. Févotte, Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. IEEE TASLP, 2010.

[2] A. Liutkus, A. Ozerov, R. Badeau, G. Richard, Spatial coding-based informed source separation. Proc. of EUSIPCO, 2012.