



Audio Inpainting, Source Separation, Audio Compression. All with a Unified Framework Based on NTF Model

Cagdas Bilen, Alexey Ozerov, Patrick Pérez

► **To cite this version:**

Cagdas Bilen, Alexey Ozerov, Patrick Pérez. Audio Inpainting, Source Separation, Audio Compression. All with a Unified Framework Based on NTF Model. MissData 2015, Jun 2015, Rennes, France. hal-01171843

HAL Id: hal-01171843

<https://hal.inria.fr/hal-01171843>

Submitted on 8 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Audio Inpainting, Source Separation, Audio Compression All with a Unified Framework Based on NTF Model

Çağdaş Bilen*, Alexey Ozerov* and Patrick Pérez

Technicolor
cagdas.bilen, alexey.ozarov, patrick.perez@technicolor.com

Keywords: Audio Inpainting, Source Separation, Informed Source Separation (ISS), Compressive Sampling, Nonnegative Tensor Factorization (NTF)

We propose a general signal recovery algorithm, that can recover J audio source signals in time, $s_{jt}^\tau, j \in \llbracket 1, J \rrbracket, t \in \llbracket 1, T \rrbracket$, given the quantized source samples, y_{jt}^τ , each observed on the support Ω_j^τ and the quantized mixture samples, $x_t^\tau = \sum_{j=1}^J s_{jt}^\tau + a_t^\tau$, observed on Ξ^τ with quantization noise a_t^τ . The sources are modelled in the short time Fourier transform (STFT) domain with a normal distribution ($s_{jfn} \sim \mathcal{N}_c(0, v_{jfn})$) where the variance tensor $\mathbf{V} = [v_{jfn}]_{j,f,n}$ has the following low-rank non-negative tensor factorization (NTF) structure [7], $v_{jfn} = \sum_{k=1}^K q_{jk} w_{fk} h_{nk}$. This model is parametrized by $\theta = \{\mathbf{Q}, \mathbf{W}, \mathbf{H}\}$, with $\mathbf{Q} = [q_{jk}]_{j,k} \in \mathbb{R}_+^{J \times K}$, $\mathbf{W} = [w_{fk}]_{f,k} \in \mathbb{R}_+^{F \times K}$ and $\mathbf{H} = [h_{nk}]_{n,k} \in \mathbb{R}_+^{N \times K}$.

We propose to recover the source signals with a generalized expectation-maximization (GEM) algorithm [4] based on multiplicative update (MU) rules [5]. The Algorithm is briefly described in Algorithm 1. Using the proposed approach, it is possible to solve a number of existing and new problems in audio signal processing:

- **Audio inpainting:** It is possible to recover arbitrary time domain losses in audio signals for applications such as signal declipping. NTF

model is used for the *first time* for the recovery of arbitrary time domain losses [3].

- **Joint audio inpainting and source separation:** It is possible to jointly perform audio inpainting and source separation to improve the performance of both tasks. Audio inpainting and source separation are performed jointly for the *first time* [1].
- **Compressed sampling-based informed source separation:** It is possible to recover the sources from their random samples and the mixture via compressive sampling-based informed source separation. This new ISS scheme uses a simple encoder that has properties of distributed coding [8, 6] and it competes with traditional ISS. The concept of distributed coding and of compressive sampling based scheme is introduced for the *first time* in the informed source separation problem [2].

The presentation and the poster will include various new results for the proposed algorithm with comparisons to state of the art methods in the different problems discussed above.

* The first and second authors have contributed equally for this work.

This work was partially supported by ANR JCJC program MAD (ANR-14-CE27-0002).

This work is submitted for a lightning talk and a poster.

Algorithm 1 GEM algorithm for NTF model estimation. Details can be found in [1, 2, 3]

- 1: **procedure** ESTIMATESOURCES-NTF($\Xi^\tau, x^\tau, \{y_{jt}^\tau\}_{j=1}^J, \{\Omega_j^\tau\}_{j=1}^J$)
 - 2: Initialize non-negative $\mathbf{Q}, \mathbf{W}, \mathbf{H}$ randomly and apply any known prior on $\mathbf{Q}, \mathbf{W}, \mathbf{H}$
 - 3: **repeat**
 - 4: Estimate sources (s_{jfn}), given $\mathbf{Q}, \mathbf{W}, \mathbf{H}, x^\tau, \Xi^\tau, \{y_{jt}^\tau\}_{j=1}^J, \{\Omega_j^\tau\}_{j=1}^J$ with Wiener filtering
 - 5: Apply constraints in the time domain, and estimate posterior power spectra of the sources ($\tilde{\mathbf{P}}$)
 - 6: Update $\mathbf{Q}, \mathbf{W}, \mathbf{H}$ given $\tilde{\mathbf{P}}$ using MU rules
 - 7: **until** convergence criteria met
 - 8: **end procedure**
-

References

- [1] Bilen, Ç., A. Ozerov, and P. Pérez (2015a, October). Audio declipping via nonnegative matrix factorization. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.
- [2] Bilen, Ç., A. Ozerov, and P. Pérez (2015b, October). Compressive sampling-based informed source separation. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.
- [3] Bilen, Ç., A. Ozerov, and P. Pérez (2015c, August). Joint audio inpainting and source separation. In *International Conference on Latent Variable Analysis and Signal Separation*.
- [4] Dempster, A., N. Laird, and D. Rubin. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39, 1–38.
- [5] Févotte, C., N. Bertin, and J. Durrieu (2009, March). Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis. *Neural Computation* 21(3), 793–830.
- [6] Girod, B., A. Aaron, S. Rane, and D. Rebollo-Monedero (2005, January). Distributed video coding. *Proceedings of the IEEE* 93(1), 71 – 83.
- [7] Ozerov, A., C. Févotte, R. Blouet, and J.-L. Durrieu (2011, May). Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'11)*, Prague, pp. 257–260.
- [8] Xiong, Z., A. Liveris, and S. Cheng (2004, September). Distributed source coding for sensor networks. *IEEE Signal Processing Magazine* 21(5), 80–94.