Multichannel Nonnegative Tensor Factorization with Structured Constraints for User-Guided Audio Source Separation

Alexey Ozerov [1], Cédric Févotte [2], Raphaël Blouet [3] and Jean-Louis Durrieu [4]

(1) INRIA – Rennes, France, (2) CNRS LTCI; Telecom ParisTech, France, (3) Yacast, Paris, France, (4) EPFL, LTS5, Lausanne, Switzerland

SARAH project: Standardization of High-Definition Audio Remastering

Audionamix — Studios Copra — Telecom ParisTech

Introduction

Separation of professionally produced music recordings is difficult:
- Sources mixed in the same direction (e.g., vocals and bass)
- Point source assumption can fail (e.g., drums)
- The problem is very ill-posed in its general formulation

User-guided separation: some input from the user is allowed

Contributions

- **Modeling:** Extension of multichannel NMF model [1] to multichannel NTF model inspired by [2]
- **Algorithmic issues:** Introduction of a new generalized EM (GEM) algorithm based on multiplicative updates (GEM-MU) that is faster than the GEM algorithm previously proposed in [1]
- **Application:** User-guided separation via specification of user-defined structured constraints (inspired by [3], where it was done for single channel case)

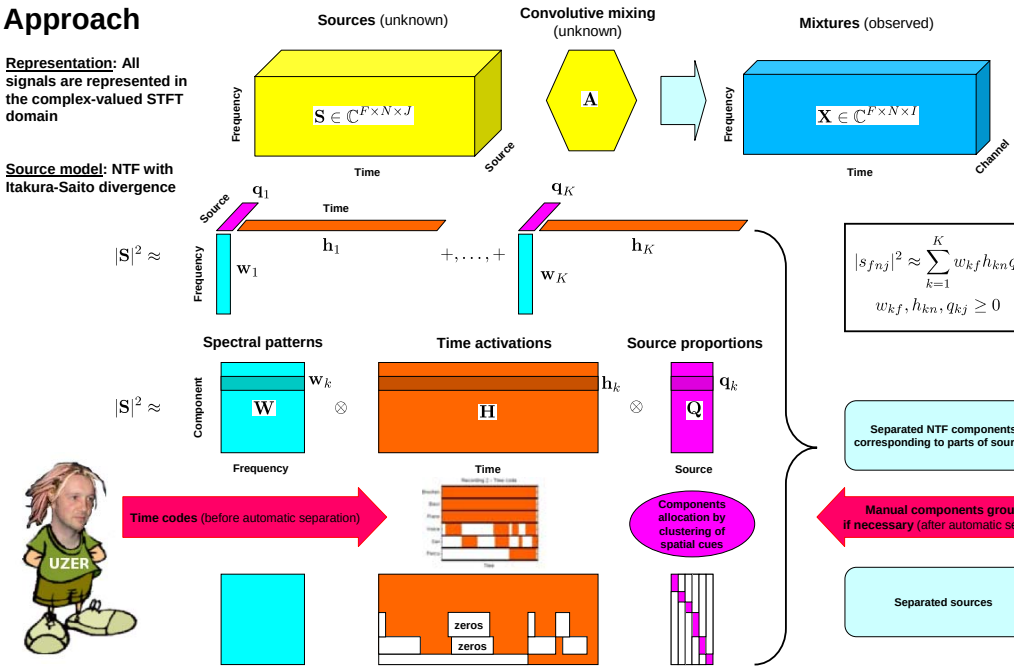
References

[1] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE TASLP*, 2010.
[2] A. Liutkus, R. Badeau, and G. Richard, "Informed source separation using latent components," in *LVA/ICA'10*, St Malo, France, 2010.
[3] H. Laurberg, M. N. Schmidt, M. G. Christensen, and S. H. Jensen, "Structured non-negative matrix factorization with sparsity patterns," in *Asilomar CSSC*, 2008.

Approach

Representation: All signals are represented in the complex-valued STFT domain

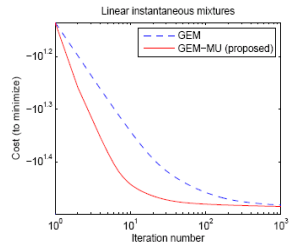
Source model: NTF with Itakura-Saito divergence



$$|s_{fnj}|^2 \approx \sum_{k=1}^{K} w_{kf} h_{kn} q_{kj}$$
$$w_{kf}, h_{kn}, q_{kj} \geq 0$$

Results

Convergence



2010 Signal Separation Evaluation Campaign (SiSEC), "Professionally produced music recordings" task

		Glen Philips "The Spirit of Shackleton"			Nine Inch Nails "The Good Soldier"		Shannon Hurley "Sunrise"				Average
		vocals	drums	bass	vocals	drums	vocals	drums	bass	piano	
Algorithm 1	SDR (dB)	3.3	2.3	-4.0	1.1	5.7	2.2	3.6	2.6	-2.3	1.6
Proposed	OPS (0-100)	19.5	31.9	14.3	30.2	27.9	15.5	39.2	8.3	18.1	22.7
Algorithm 2	SDR (dB)	-0.3	-	-	-2.6	-	0.8	-	-	-	-
J. Janer & R. Marxer [4]	OPS (0-100)	15.9	-	-	18.8	-	15.2	-	-	-	-
Algorithm 3	SDR (dB)	3.9	3.6	-2.0	1.1	1.2	2.2	4.7	3.4	-3.8	1.6
M. Spiertz [10]	OPS (0-100)	15.4	37.3	8.7	25.2	25.0	8.0	40.6	5.8	10.4	19.6
STFT	SDR (dB)	5.6	6.4	2.0	1.5	5.1	7.8	7.3	8.2	0.7	4.9
Ideal Binary Mask	OPS (0-100)	21.0	30.6	11.3	29.3	37.9	15.4	35.5	20.6	19.2	24.5
Cochleagram	SDR (dB)	3.9	1.1	0.7	1.4	1.6	6.1	1.5	1.6	0.4	2.0
Ideal Binary Mask	OPS (0-100)	15.7	26.2	15.6	17.6	37.5	11.2	42.8	30.4	12.1	23.2

Evaluation measures: Signal to Distortion Ratio (SDR) and Overall Perceptual Score (OPS)

Conclusion

- Novel user-guided audio source separation method based on a multichannel NTF model with structured constraints
- The approach allows high quality source separation of real recordings using limited user input