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

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

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