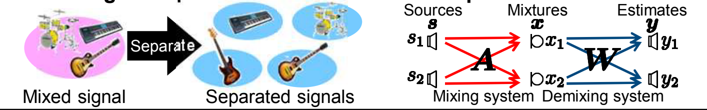
EFFICIENT MULTICHANNEL NONNEGATIVE MATRIX FACTORIZATION EXPLOITING RANK-1 SPATIAL MODEL

National University SOKENDAI The Graduate University for Advanced Studies
NII 国立情報学研究所 National Institute of Informatics
NTT 東京大学 THE UNIVERSITY OF TOKYO

Daichi Kitamura (SOKENDAI), Nobutaka Ono (NII/SOKENDAI), Hiroshi Sawada (NTT), Hirokazu Kameoka (The Univ. of Tokyo/NTT), Hiroshi Saruwatari (The Univ. of Tokyo)

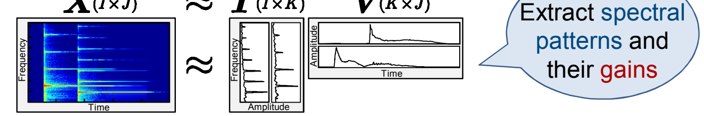
1. Abstract

- **Music signal separation and blind source separation**



- **Nonnegative matrix factorization (NMF)**

Observed spectrogram $X(I \times J)$ ≈ Basis matrix $T(I \times K)$ Activation matrix $V(K \times J)$



Problems of conventional methods

Independent vector analysis (IVA): unsuitable for music signals
Multichannel NMF (MNMF): computational costs and strong dependence for initialization

MNMF exploiting rank-1 spatial model

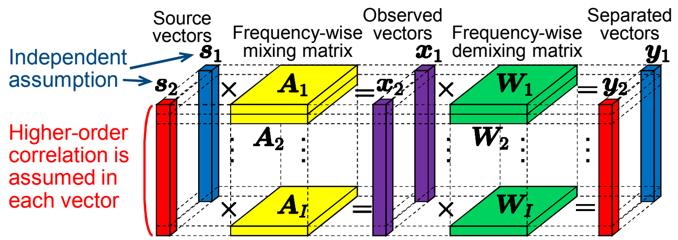
For multichannel music separation task, we propose an **efficient algorithm of MNMF** (fast convergence and good performance) that exploits **rank-1 spatial model**.

We propose a new MNMF estimating demixing matrix W.
Optimization: Alternative iteration of IVA and NMF update rules
Experiments show the efficiency of the proposed method.

2. Conventional methods

• Independent vector analysis (IVA)

Multivariate extension of frequency-domain ICA
Independence between frequency vectors is utilized for estimating W_i.
Permutation problem can be solved by assuming multivariate prior.



Demixing filters W_i are estimated so that y_1 and y_2 are independent.

Cost function of IVA (prior: spherical Laplacian distribution)

$$\mathcal{J}_{\mathrm{IVA}} = \sum_m \frac{1}{J} \sum_j \| y_{j,m} \|_2 - \sum_i \log |\det W_i|$$

• Multichannel NMF (MNMF)

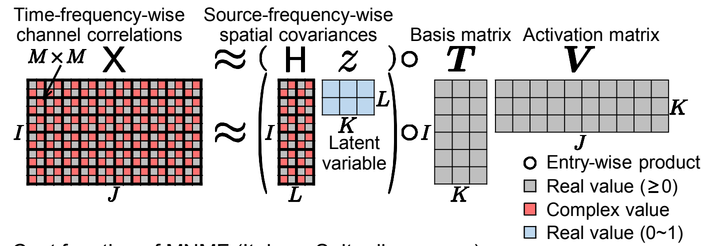
Cross-correlations between channels (spatial covariances) for each source and frequency are introduced.
Sources are separated based on the clustering of spatial covariance.

Multichannel vector

$$x_{ij} = \begin{pmatrix} x_{ij,1} \\ \vdots \\ x_{ij,M} \end{pmatrix}$$

Instantaneous covariance matrix ($M \times M$)

$$\mathrm{X}_{ij} = x_{ij} x_{ij}^{\mathrm{H}} = \begin{pmatrix} |x_{ij,1}|^2 & \cdots & x_{ij,1} x_{ij,M}^* \\ \vdots & \ddots & \vdots \\ x_{ij,M} x_{ij,1}^* & \cdots & |x_{ij,M}|^2 \end{pmatrix}$$



Cost function of MNMF (Itakura-Saito divergence)

$$\mathcal{J}_{\mathrm{MNMF}} = \sum_{i,j} \left[\mathrm{tr} \left(\mathrm{X}_{ij} \left(\sum_k \left(\sum_m \mathrm{H}_{i,m} z_{mk} \right) t_{ik} v_{kj} \right)^{-1} \right) + \log \det \sum_k \left(\sum_m \mathrm{H}_{i,m} z_{mk} \right) t_{ik} v_{kj} \right]$$

3. Proposed method

• Motivation

IVA estimates frequency-wise **demixing filter W_i**.
➡ Fast convergence, strict source model
MNMF estimates source model TV and **mixing system $\mathrm{H}_{i,m}$**.
➡ Much computational time, strong dependence for initialization

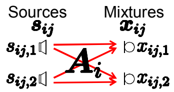
Proposed MNMF estimates source model TV and demixing filter W_i.
➡ Fast convergence, flexible source model

• New MNMF estimating demixing matrix

Model all the spatial covariance matrices $\mathrm{H}_{i,m}$ as **rank-1 matrix**

Rank-1 spatial model

- The mixing system is approximated as instantaneous time-invariant mixing matrix A_i.



Rank-1 spatial model: $\mathrm{H}_{i,m} = a_{i,m} a_{i,m}^{\mathrm{H}}$, $A_i = (a_{i,1} \ \cdots \ a_{i,M})$

Substitute / Substitute

$$\mathcal{J}_{\mathrm{MNMF}} = \sum_{i,j} \left[\mathrm{tr} \left(\mathrm{X}_{ij} \left(\sum_k \left(\sum_m \mathrm{H}_{i,m} z_{mk} \right) t_{ik} v_{kj} \right)^{-1} \right) + \log \det \sum_k \left(\sum_m \mathrm{H}_{i,m} z_{mk} \right) t_{ik} v_{kj} \right]$$

Reform the cost function using the mixing matrix A_i
Transform the variables by $W_i = A_i^{-1}$, $y_{ij} = W_i x_{ij}$

$$\mathcal{J} = \sum_{i,j} \left[\sum_m \frac{|y_{ij,m}|^2}{\sum_k z_{mk} t_{ik} v_{kj}} - 2 \log |\det W_i| + \sum_m \log \sum_k z_{mk} t_{ik} v_{kj} \right]$$

• Optimization of proposed MNMF

The cost function consists of IVA and NMF terms.

$$\mathcal{J} = \sum_{i,j} \left[\sum_m \frac{|y_{ij,m}|^2}{\sum_k z_{mk} t_{ik} v_{kj}} - 2 \log |\det W_i| + \sum_m \log \sum_k z_{mk} t_{ik} v_{kj} \right]$$

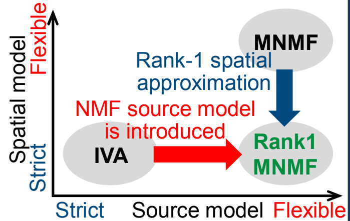
IVA
Single-channel NMF

$$\mathcal{J}_{\mathrm{NMF}} = \sum_{i,j} \left[\frac{|y_{ij}|^2}{\sum_k t_{ik} v_{kj}} + \log \sum_k t_{ik} v_{kj} \right]$$

All the variables W, Z, T, and V are optimized by the alternative iteration of IVA and simple NMF update rules.
➡ Fast update rules have been proposed for both IVA and NMF.

The proposed method is an intermediate model between IVA and conventional MNMF.

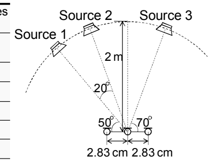
MNMF with rank-1 model is equivalent to IVA with NMF source model.



4. Experiment

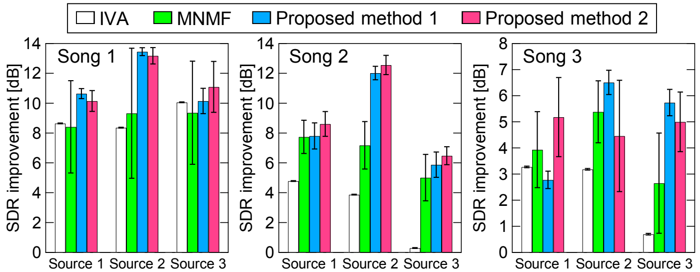
• Conditions

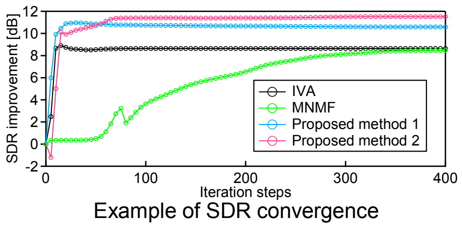
Source signal	SiSEC music signals convoluted with RWCP impulse responses 3 sources and 3 channels
Comparison	IVA, MNMF, Proposed method 1 (w/o latent variable), Proposed method 2 (with latent variable)
FFT length	8192 (512 ms)
Window length	2048 (128 ms, Han window)
# of bases	Proposed method 1: 30 per source
	Proposed method 2: 90
# of iteration	200
# of trials	10 trials with different initial values
Evaluation criterion	Average SDR improvement and its deviation



Impulse response (Rev. time 300 ms)

• Results





Example of SDR convergence

Example of computational times

IVA	91.6 s
MNMF	4498.5 s
Prop. 1	121.0 s
Prop. 2	173.4 s

Sound demo ➡