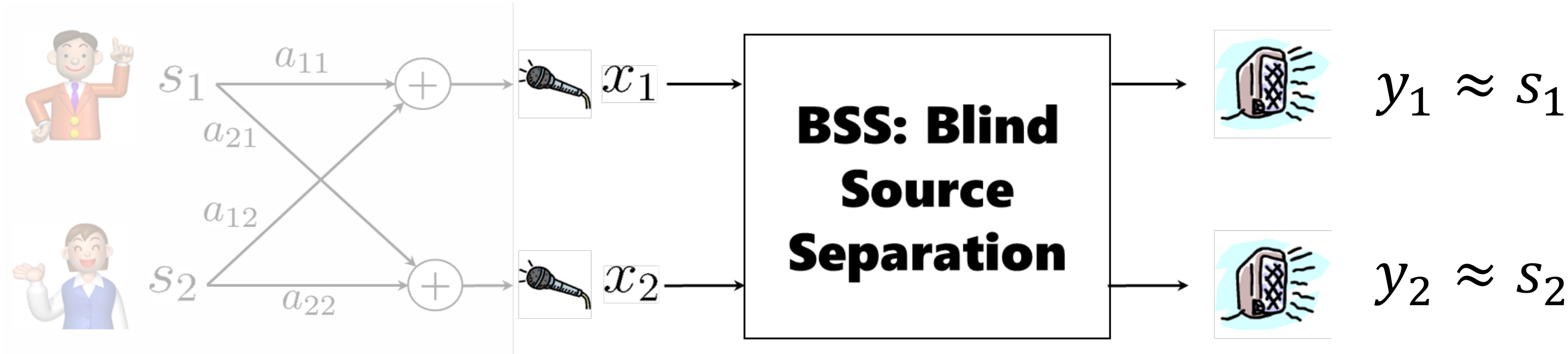
Presented in **IEEE ICASSP 2022** AUD-16.1

Multi-frame Full-rank Spatial Covariance Analysis for Underdetermined BSS in Reverberant Environment

Hiroshi Sawada, Rintaro Ikeshita,
Keisuke Kinoshita, Tomohiro Nakatani

NTT Communication Science Laboratories,
NTT Corporation

Background



	overdetermined $N \leq M$	underdetermined $N > M$
low reverberant	ICA	FCA
high reverberant	WPE	

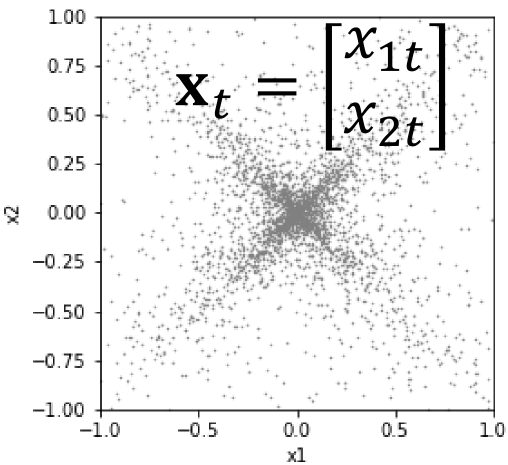
- ICA: Independent Component Analysis
- FCA: Full-rank spatial Covariance Analysis
- WPE: Weighted Prediction Error

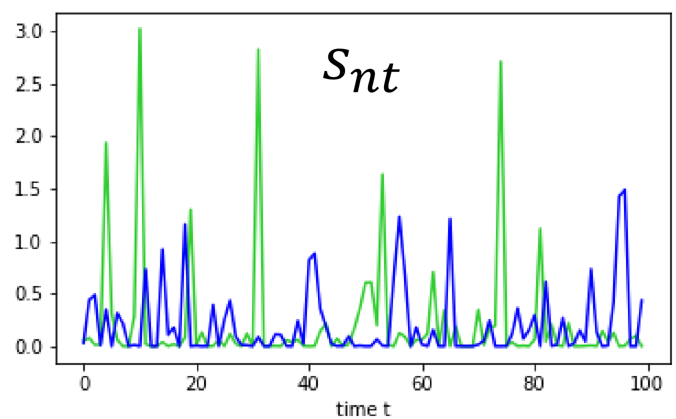
ICA, FCA: BSS methods

WPE: a blind dereverberation method

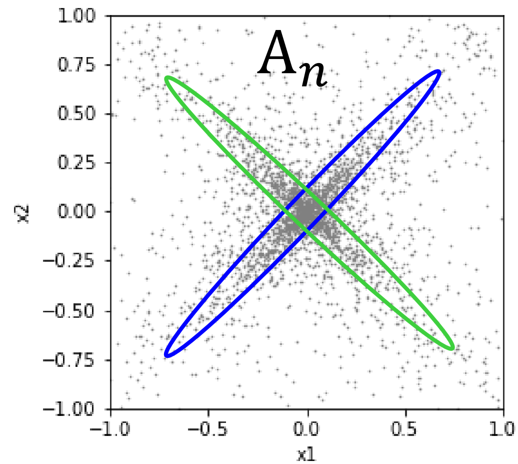
FCA model

- **Observation vector**
  - sum of source components $\mathbf{x}_t = \sum_{n=1}^{N} \mathbf{c}_{nt}$ $\quad \mathbf{c}_{nt} = \begin{bmatrix} c_{1nt} \\ \vdots \\ c_{Mnt} \end{bmatrix} \in \mathbb{C}^M$
- **Source component vector**
  - follows a zero-mean Gaussian distribution with a covariance matrix C_{nt} $\quad p(\mathbf{c}_{nt}) = \mathcal{N}(\mathbf{c}_{nt} \mid \mathbf{0}, \mathsf{C}_{nt})$ $\quad \mathsf{C}_{nt} = \mathsf{s}_{nt}\mathsf{A}_n$
- **Parameters** $\theta = \{\{\{\mathsf{s}_{nt}\}_{t=1}^{T}, \mathsf{A}_n\}_{n=1}^{N}\}$





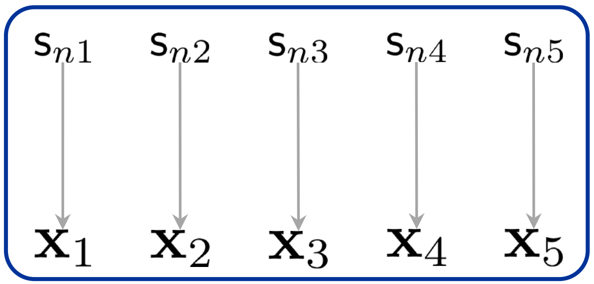
Temporal power of source n

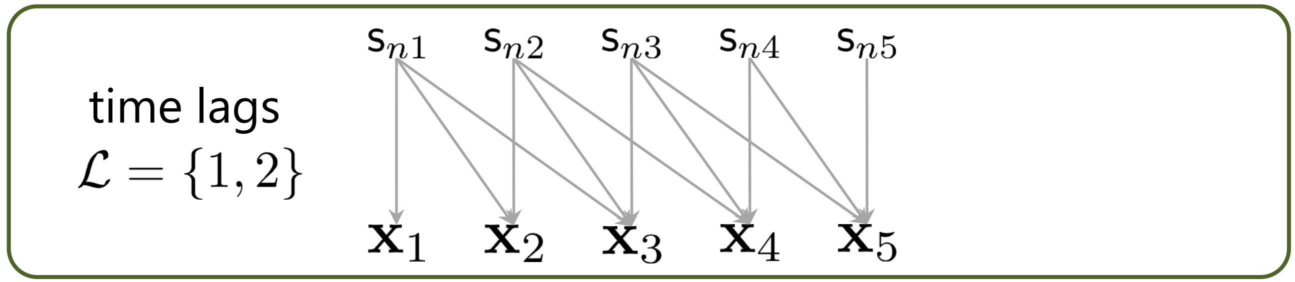


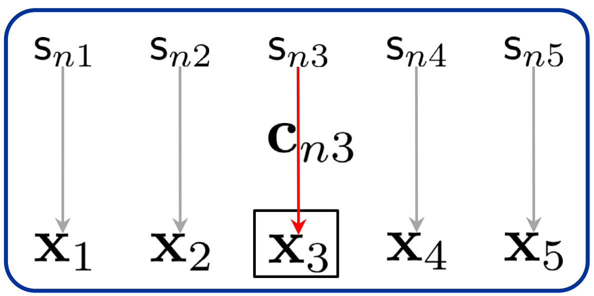
Spatial property from source n to all microphones

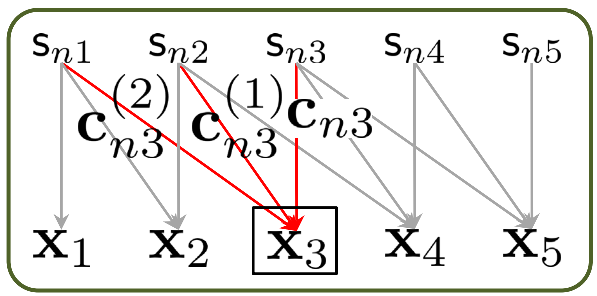
FCA and its Extensions

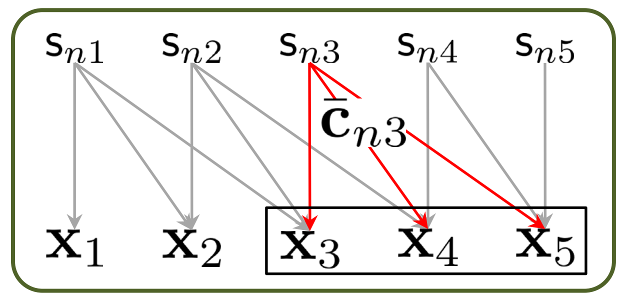
Original	Conventional	Our proposal
FCA	FCAd	mfFCA
	delayed source components	multi-frame











A_n

$\mathsf{A}_n\,,\ \mathsf{A}_n^{(1)}\,,\ \mathsf{A}_n^{(2)}$

$$\begin{bmatrix} \mathsf{A}_n & \mathbf{A}_n^{(0,1)} & \mathbf{A}_n^{(0,2)} \\ \mathbf{A}_n^{(1,0)} & \mathsf{A}_n^{(1)} & \mathbf{A}_n^{(1,2)} \\ \mathbf{A}_n^{(2,0)} & \mathbf{A}_n^{(2,1)} & \mathsf{A}_n^{(2)} \end{bmatrix}$$

Proposed mfFCA model

- **Multi-frame vectors**

observation $\bar{\mathbf{x}}_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \end{bmatrix}$

source component $\bar{\mathbf{c}}_{nt} = \begin{bmatrix} \mathbf{c}_{nt} \\ \mathbf{c}^{(1)}_{n(t+1)} \\ \mathbf{c}^{(2)}_{n(t+2)} \end{bmatrix}$

- **Source component vector $\bar{\mathbf{c}}_{nt}$**
  - zero-mean Gaussian distribution $p(\bar{\mathbf{c}}_{nt}) = \mathcal{N}(\bar{\mathbf{c}}_{nt} \mid \mathbf{0}, \bar{\mathsf{C}}_{nt})$
  - covariance matrix has larger dimensionality

$$\bar{\mathsf{C}}_{nt} = \mathsf{s}_{nt}\bar{\mathsf{A}}_n \qquad \bar{\mathsf{A}}_n = \begin{bmatrix} \mathsf{A}_n & \mathbf{A}_n^{(0,1)} & \mathbf{A}_n^{(0,2)} \\ \mathbf{A}_n^{(1,0)} & \mathsf{A}_n^{(1)} & \mathbf{A}_n^{(1,2)} \\ \mathbf{A}_n^{(2,0)} & \mathbf{A}_n^{(2,1)} & \mathsf{A}_n^{(2)} \end{bmatrix}$$

- **Observation vector $\bar{\mathbf{x}}_t$**
  - zero-mean Gaussian distribution $p(\bar{\mathbf{x}}_t \mid \theta) = \mathcal{N}(\bar{\mathbf{x}}_t \mid \mathbf{0}, \bar{\mathsf{X}}_t)$
  - covariance matrix has additional terms specific to mfFCA

$$\bar{\mathsf{X}}_t = \underbrace{\underbrace{\begin{bmatrix} \mathsf{X}_t & & \\ & \ddots & \\ & & \mathsf{X}_{t+l_L} \end{bmatrix}}_{\text{FCAd}} + \sum_{n=1}^{N} \mathrm{BoffDiag}\,\bar{\mathsf{C}}_{nt}}_{\text{mfFCA}}$$

mfFCA: EM algorithm

- **For optimizing parameters** $\theta = \{\{\{\mathsf{s}_{nt}\}_{t=1}^{T}, \bar{\mathsf{A}}_n\}_{n=1}^{N}\}$
- **E-step**
  - conditional distribution $p(\bar{\mathbf{c}}_{nt} \mid \bar{\mathbf{x}}_t, \theta) = \mathcal{N}(\bar{\mathbf{c}}_{nt} \mid \boldsymbol{\mu}_{nt}^{(\bar{c})}, \Sigma_{nt}^{(\bar{c})})$

    mean vector $\boldsymbol{\mu}_{nt}^{(\bar{c})} = \bar{\mathsf{C}}_{nt}\bar{\mathsf{X}}_t^{-1}\bar{\mathbf{x}}_t$

    covariance matrix $\Sigma_{nt}^{(\bar{c})} = \bar{\mathsf{C}}_{nt} - \bar{\mathsf{C}}_{nt}\bar{\mathsf{X}}_t^{-1}\bar{\mathsf{C}}_{nt}$
- **M-step**
  - optimize parameters

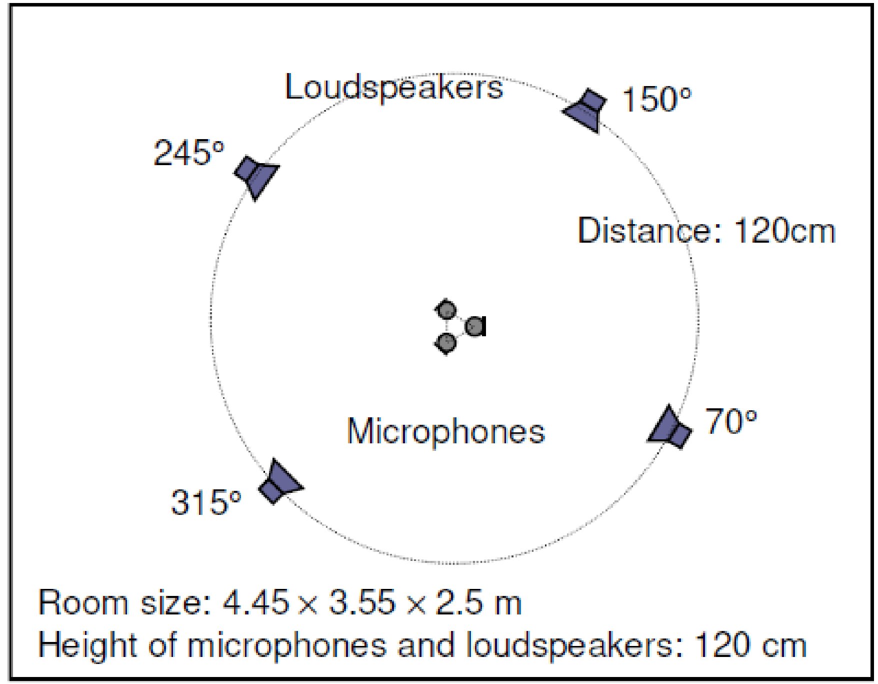
$$\bar{\mathsf{A}}_n \leftarrow \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\mathsf{s}_{nt}}\widetilde{\mathsf{C}}_{nt}$$

with $\widetilde{\mathsf{C}}_{nt} = \boldsymbol{\mu}_{nt}^{(\bar{c})}\boldsymbol{\mu}_{nt}^{(\bar{c})\mathsf{H}} + \Sigma_{nt}^{(\bar{c})}$

$$\mathsf{s}_{nt} \leftarrow \frac{1}{M(L+1)}\mathrm{tr}\left(\bar{\mathsf{A}}_n^{-1}\widetilde{\mathsf{C}}_{nt}\right)$$

Experiments

- **Conditions**
  - $M = 3$ microphones
  - $N = 4$ sources
  - 6-second speeches
  - reverberation time: 130 ms to 450 ms



- **Separation performances**
  - measured in signal-to-distortion ratios (SDRs)
  - did not aim for dereverberation (we used source images with reverberations at microphones as reference signals)

Overall results

5 methods

FCA

original, baseline

FCAd

conventional

$\mathcal{L} = \{2\}$

$\mathcal{L} = \{2, 4\}$

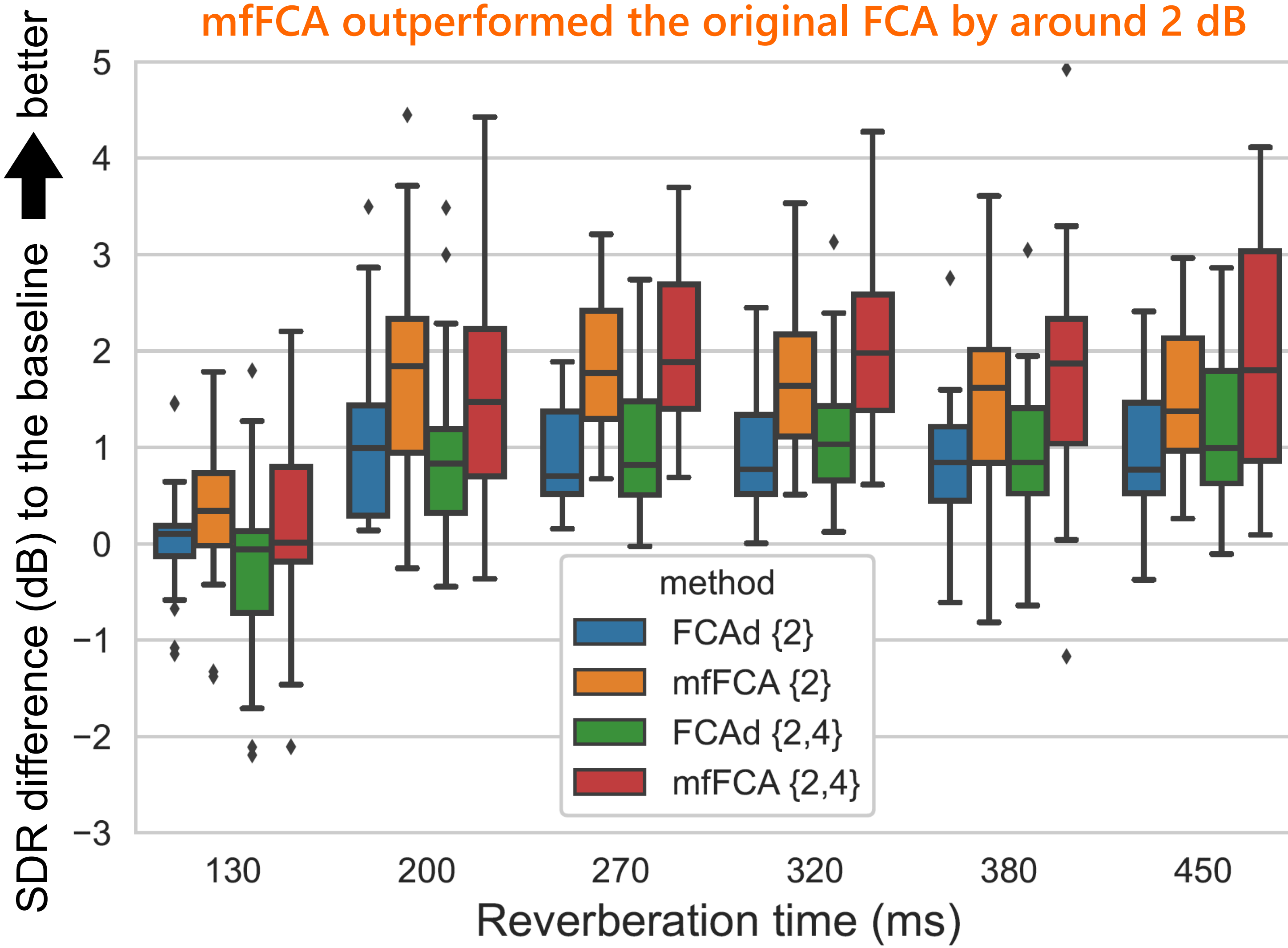
mfFCA

our proposal

$\mathcal{L} = \{2\}$

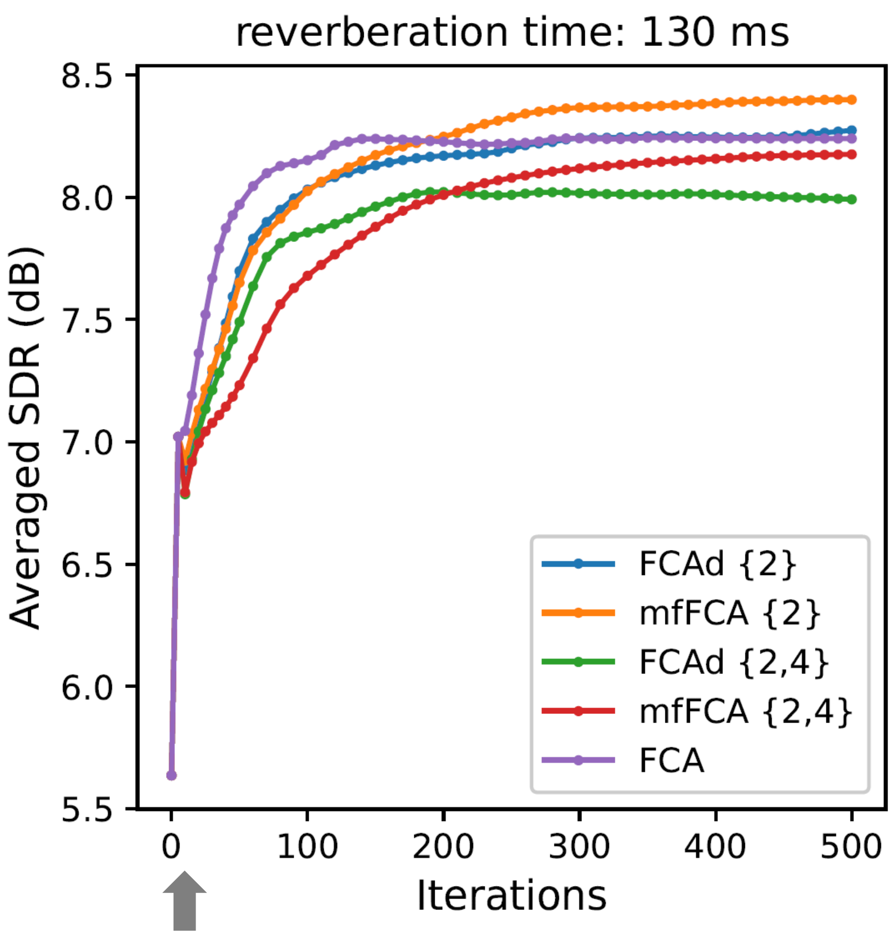
$\mathcal{L} = \{2, 4\}$

mfFCA outperformed the original FCA by around 2 dB

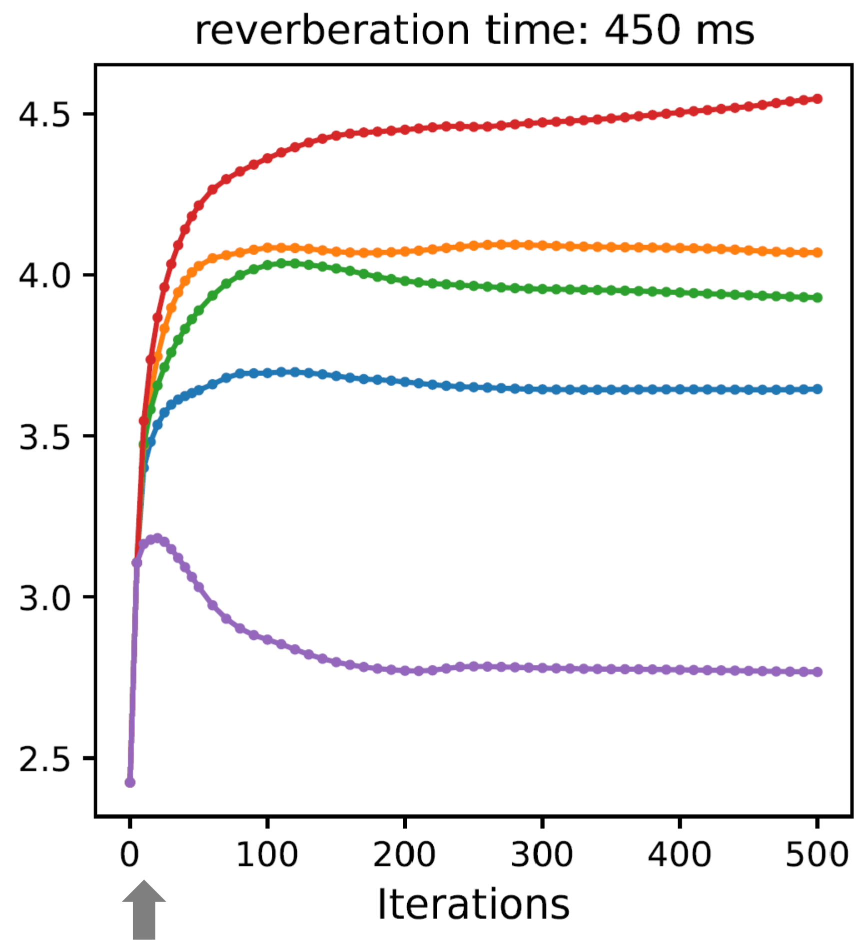


Convergence behavior

Low reverberant case



High reverberant case



- FCA parameters were initialized by the procedure shown in [29].
- The first 5 iterations were by the original FCA model and updates.

Conclusion

- **A new FCA model** mfFCA
  - source components span multiple time frames
  - modeled with covariance matrix of larger dimensionality

$$p(\bar{\mathbf{c}}_{nt}) = \mathcal{N}(\bar{\mathbf{c}}_{nt} \mid \mathbf{0}, \bar{\mathsf{C}}_{nt})$$

$$\bar{\mathbf{c}}_{nt} = \begin{bmatrix} \mathbf{c}_{nt} \\ \mathbf{c}^{(1)}_{n(t+1)} \\ \mathbf{c}^{(2)}_{n(t+2)} \end{bmatrix}$$

$$\bar{\mathsf{C}}_{nt} = \mathsf{s}_{nt}\bar{\mathsf{A}}_n \qquad \bar{\mathsf{A}}_n = \begin{bmatrix} \mathsf{A}_n & \mathbf{A}_n^{(0,1)} & \mathbf{A}_n^{(0,2)} \\ \mathbf{A}_n^{(1,0)} & \mathsf{A}_n^{(1)} & \mathbf{A}_n^{(1,2)} \\ \mathbf{A}_n^{(2,0)} & \mathbf{A}_n^{(2,1)} & \mathsf{A}_n^{(2)} \end{bmatrix}$$

- **Developed**
  - the whole probabilistic models and EM algorithm
- **Experimental results**
  - show that the proposed method considerably improved the separation performance for underdetermined reverberant convolutive mixtures
- **Future work**
  - evaluating the dereverberation capability of mfFCA
  - reducing the computational complexity further (we have already accelerated the algorithm computation by a GPU)