SS-L6, ICASSP 2019, Brighton, UK
(May 16th, 2019)

Similarity Search-based Blind Source Separation

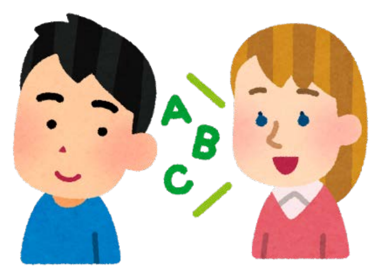
Hiroshi Sawada Kazuo Aoyama

NTT Communication Science Laboratories,
NTT Corporation

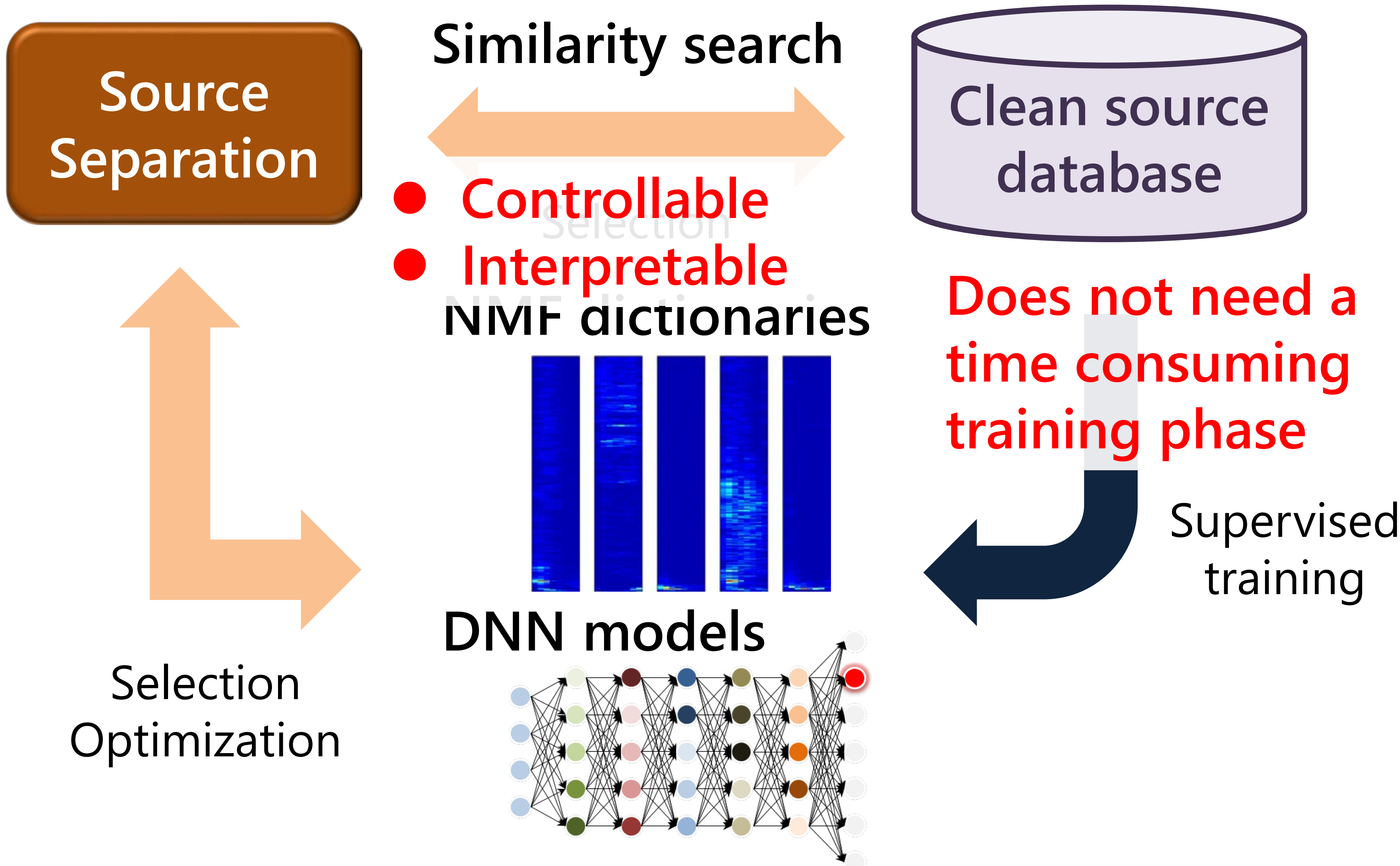
Background & Motivation

- **BSS methods**
  - Independence: ICA, IVA
  - Low-Rankness: ILRMA
  - Needs enough amount of observations ($\geq$ 3 sec.)
- **Time-varying Environments**
  - Short observations ($\leq$ 2 sec.)

- **Similarity search** on a clean source database
  - Human can separate mixtures if there is something familiar to us in the mixtures



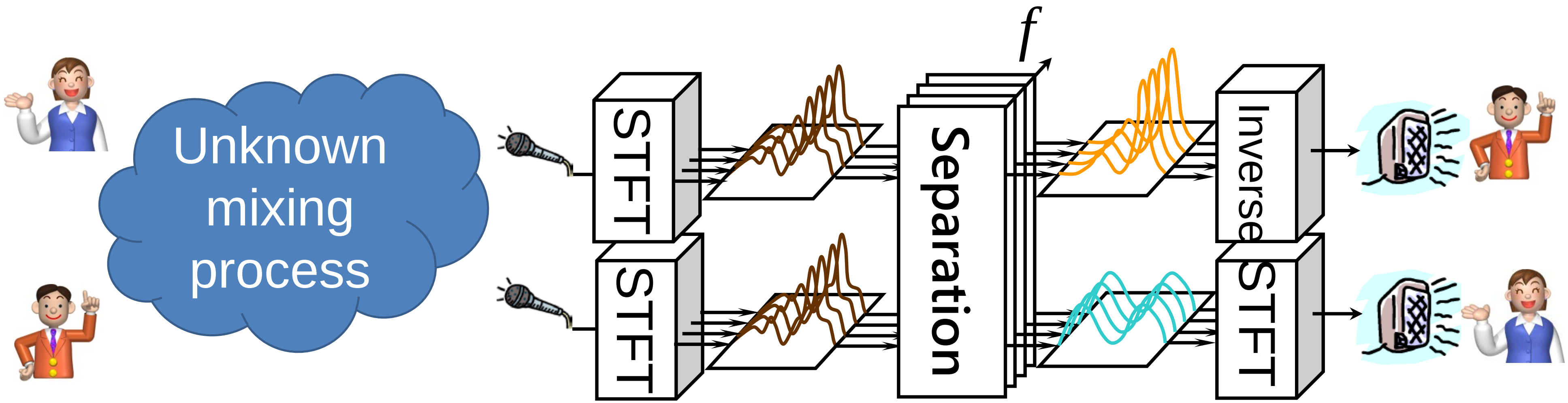
Supervised learning?



Outline

1. **Existing BSS methods**
   - Frequency-domain BSS
   - IVA, ICA, ILRMA
2. **Proposed method**
   - **SSBSS**: Similarity Search-based BSS
   - Differs in variance parameter updates
3. **Experiments**
   - Clean source databases: close and open
   - Convergence behavior
   - Computational time with a GPU

Frequency-domain BSS



Separation

$$\mathbf{y}_{ft} = \mathbf{W}_f \mathbf{x}_{ft} \qquad f = 1, \ldots, F$$

$$\mathbf{y}_{ft} = \begin{bmatrix} y_{ft1} \\ y_{ft2} \end{bmatrix} \qquad \mathbf{W}_f = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \qquad \mathbf{x}_{ft} = \begin{bmatrix} x_{ft1} \\ x_{ft2} \end{bmatrix}$$

Separated signal — Separation matrix — Mixture

Objective function

$$\mathcal{J}(\{\mathbf{W}_f\}_{f=1}^F) = \sum_{t=1}^{T}\sum_{n=1}^{N} G(\tilde{\mathbf{y}}_{tn}) - 2T\sum_{f=1}^{F}\log|\det \mathbf{W}_f|$$

$f = 1, \ldots, F$ Frequency bins

$n = 1, \ldots, N$ Separated signals

$t = 1, \ldots, T$ Time frames

Contrast function G

Local Gaussian Model

$$p(\tilde{\mathbf{y}}_{tn}) = \prod_{f=1}^{F} \frac{1}{\pi v_{tn}} \exp\left(-\frac{|y_{ftn}|^2}{v_{tn}}\right)$$

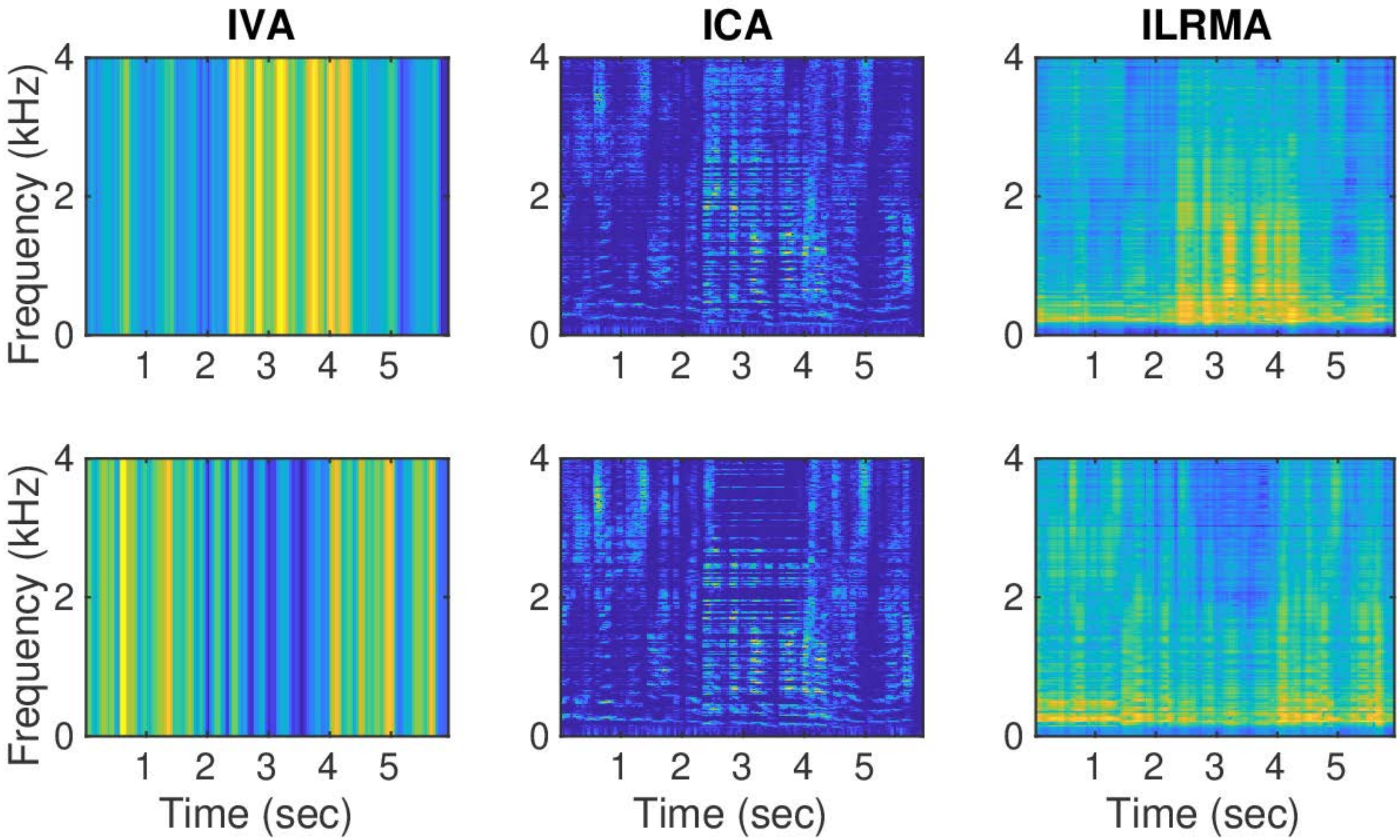
variance parameter v

IVA $$G(\tilde{\mathbf{y}}_{tn}) = \sum_{f=1}^{F}\left(\frac{|y_{ftn}|^2}{v_{tn}} + \log v_{tn}\right)$$

ICA $$G(\tilde{\mathbf{y}}_{tn}) = \sum_{f=1}^{F}\left(\frac{|y_{ftn}|^2}{v_{ftn}} + \log v_{ftn}\right)$$

ILRMA $$G(\tilde{\mathbf{y}}_{tn}) = \sum_{f=1}^{F}\left(\frac{|y_{ftn}|^2}{v_{ftn}} + \log v_{ftn}\right), \quad v_{ftn} = \sum_{k=1}^{K} b_{fnk}a_{tnk}$$

Variance parameters v



- Time varying activity
- Flat spectrum

Permutation problem

Low-rank model well estimated

IVA Optimization

Variance update

$$v_{tn} \leftarrow \frac{1}{F}\sum_{f=1}^{F} |y_{ftn}|^2$$

$$\mathbf{y}_{ft} = \mathbf{W}_f \mathbf{x}_{ft}$$

Frequency-wise separation matrix update $\mathbf{W}_f$

Weighted covariance matrix

$$\mathbf{U}_{fn} = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{v_{tn}} \mathbf{x}_{ft}\mathbf{x}_{ft}^{\mathsf{H}}$$

Solve HEAD: Hybrid Exact-Approximate Diagonalization

[Yeredor 2009]

$$\mathbf{w}_{fk}^{\mathsf{H}} \mathbf{U}_{fn} \mathbf{w}_{fn} = \delta_{kn}$$

N=2 case

$$\mathbf{w}_1^{\mathsf{H}}\mathbf{U}_1\mathbf{w}_1 = 1 \qquad \mathbf{w}_1^{\mathsf{H}}\mathbf{U}_2\mathbf{w}_2 = 0$$

$$\mathbf{w}_2^{\mathsf{H}}\mathbf{U}_1\mathbf{w}_1 = 0 \qquad \mathbf{w}_2^{\mathsf{H}}\mathbf{U}_2\mathbf{w}_2 = 1$$

[Ono 2011]

$$\mathbf{w}_{fn} \leftarrow (\mathbf{W}_f \mathbf{U}_{fn})^{-1}\mathbf{e}_n$$

$$\mathbf{w}_{fn} \leftarrow \frac{\mathbf{w}_{fn}}{\sqrt{\mathbf{w}_{fn}^{\mathsf{H}}\mathbf{U}_{fn}\mathbf{w}_{fn}}}$$

$$n = 1, \ldots, N$$

ICA Optimization

Variance update

$$v_{ftn} \leftarrow |y_{ftn}|^2$$

$$\mathbf{y}_{ft} = \mathbf{W}_f \mathbf{x}_{ft}$$

Frequency-wise separation matrix update $\mathbf{W}_f$

Weighted covariance matrix

$$\mathbf{U}_{fn} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{v_{tn}}\mathbf{x}_{ft}\mathbf{x}_{ft}^{\mathsf{H}}$$

Solve HEAD: Hybrid Exact-Approximate Diagonalization [Yeredor 2009]

$$\mathbf{w}_{fk}^{\mathsf{H}}\mathbf{U}_{fn}\mathbf{w}_{fn} = \delta_{kn}$$

N=2 case

$$\mathbf{w}_1^{\mathsf{H}}\mathbf{U}_1\mathbf{w}_1 = 1 \qquad \mathbf{w}_1^{\mathsf{H}}\mathbf{U}_2\mathbf{w}_2 = 0$$

$$\mathbf{w}_2^{\mathsf{H}}\mathbf{U}_1\mathbf{w}_1 = 0 \qquad \mathbf{w}_2^{\mathsf{H}}\mathbf{U}_2\mathbf{w}_2 = 1$$

[Ono 2011]

$$\mathbf{w}_{fn} \leftarrow (\mathbf{W}_f\mathbf{U}_{fn})^{-1}\mathbf{e}_n$$

$$\mathbf{w}_{fn} \leftarrow \frac{\mathbf{w}_{fn}}{\sqrt{\mathbf{w}_{fn}^{\mathsf{H}}\mathbf{U}_{fn}\mathbf{w}_{fn}}}$$

$$n = 1, \ldots, N$$

ILRMA Optimization

Variance update

$$v_{ftn} \approx |y_{ftn}|^2$$

$$v_{ftn} \leftarrow \sum_{k=1}^{K} b_{fnk} a_{tnk}$$

$$\mathbf{y}_{ft} = \mathbf{W}_f \mathbf{x}_{ft}$$

Frequency-wise separation matrix update $\mathbf{W}_f$

Weighted covariance matrix

$$\mathbf{U}_{fn} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{v_{tn}} \mathbf{x}_{ft} \mathbf{x}_{ft}^{\mathsf{H}}$$

Solve HEAD: Hybrid Exact-Approximate Diagonalization

[Yeredor 2009]

$$\mathbf{w}_{fk}^{\mathsf{H}} \mathbf{U}_{fn} \mathbf{w}_{fn} = \delta_{kn}$$

N=2 case

$$\mathbf{w}_1^{\mathsf{H}} \mathbf{U}_1 \mathbf{w}_1 = 1 \qquad \mathbf{w}_1^{\mathsf{H}} \mathbf{U}_2 \mathbf{w}_2 = 0$$

$$\mathbf{w}_2^{\mathsf{H}} \mathbf{U}_1 \mathbf{w}_1 = 0 \qquad \mathbf{w}_2^{\mathsf{H}} \mathbf{U}_2 \mathbf{w}_2 = 1$$

[Ono 2011]

$$\mathbf{w}_{fn} \leftarrow (\mathbf{W}_f \mathbf{U}_{fn})^{-1} \mathbf{e}_n$$

$$\mathbf{w}_{fn} \leftarrow \frac{\mathbf{w}_{fn}}{\sqrt{\mathbf{w}_{fn}^{\mathsf{H}} \mathbf{U}_{fn} \mathbf{w}_{fn}}}$$

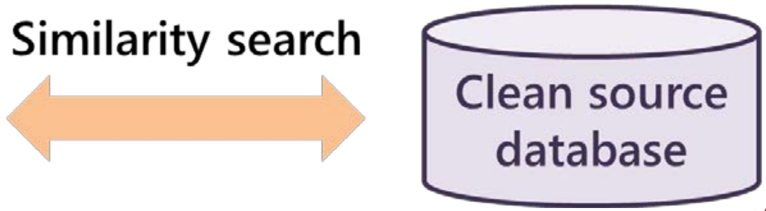
$$n = 1, \ldots, N$$

Outline

1. **Existing BSS methods**
   - Frequency-domain BSS
   - IVA, ICA, ILRMA
2. **Proposed method**
   - **SSBSS**: Similarity Search-based BSS
   - Differs in variance parameter updates
3. **Experiments**
   - Clean source databases: close and open
   - Convergence behavior
   - Computational time with a GPU

SSBSS Optimization

Variance update



$$\mathbf{y}_{ft} = \mathbf{W}_f \mathbf{x}_{ft}$$

Frequency-wise separation matrix update $\mathbf{W}_f$

Weighted covariance matrix

$$\mathbf{U}_{fn} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{v_{tn}}\mathbf{x}_{ft}\mathbf{x}_{ft}^{\mathsf{H}}$$

Solve HEAD: Hybrid Exact-Approximate Diagonalization

[Yeredor 2009]

$$\mathbf{w}_{fk}^{\mathsf{H}}\mathbf{U}_{fn}\mathbf{w}_{fn} = \delta_{kn}$$

N=2 case

$$\mathbf{w}_1^{\mathsf{H}}\mathbf{U}_1\mathbf{w}_1 = 1 \qquad \mathbf{w}_1^{\mathsf{H}}\mathbf{U}_2\mathbf{w}_2 = 0$$

$$\mathbf{w}_2^{\mathsf{H}}\mathbf{U}_1\mathbf{w}_1 = 0 \qquad \mathbf{w}_2^{\mathsf{H}}\mathbf{U}_2\mathbf{w}_2 = 1$$

[Ono 2011]

$$\mathbf{w}_{fn} \leftarrow (\mathbf{W}_f\mathbf{U}_{fn})^{-1}\mathbf{e}_n$$

$$\mathbf{w}_{fn} \leftarrow \frac{\mathbf{w}_{fn}}{\sqrt{\mathbf{w}_{fn}^{\mathsf{H}}\mathbf{U}_{fn}\mathbf{w}_{fn}}}$$

$$n = 1, \ldots, N$$

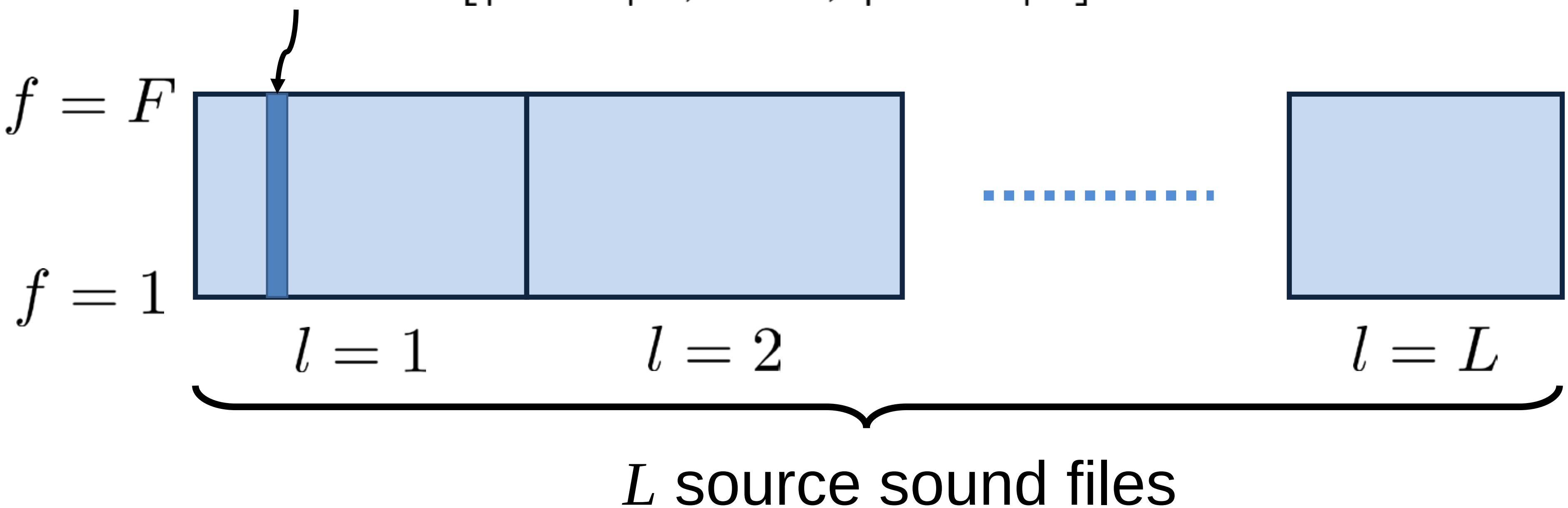
Clean source database

- **Database**

$$\mathbb{S} = \{\{\breve{\mathbf{s}}_{tl}\}_{t=1}^{T_l}\}_{l=1}^{L}$$

  - **Entry:** F-dimensional power spectra vector

$$\breve{\mathbf{s}}_{tl} = [|s_{1tl}|^2, \ldots, |s_{Ftl}|^2]^\mathsf{T} \qquad s_{ftl} \in \mathbb{C}$$



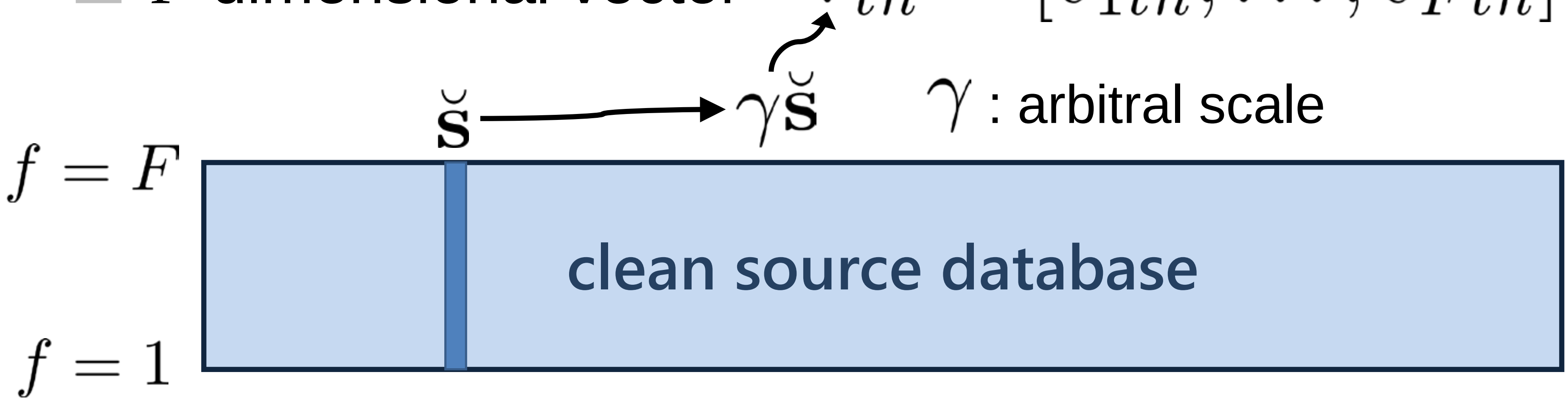
SSBSS Objective function

- **Objective function (same structure)**

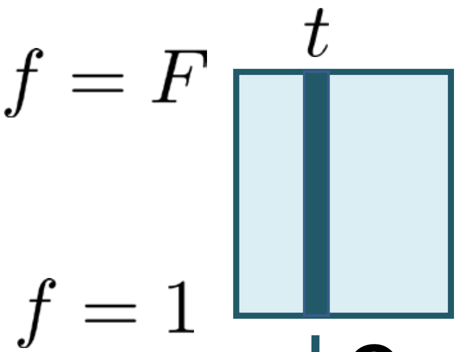
$$\mathcal{J}(\{\mathbf{W}_f\}_{f=1}^{F}) = \sum_{t=1}^{T}\sum_{n=1}^{N} G(\tilde{\mathbf{y}}_{tn}) - 2T\sum_{f=1}^{F}\log|\det \mathbf{W}_f|$$

Contrast function G $\quad G(\tilde{\mathbf{y}}_{tn}) = \sum_{f=1}^{F}\left(\frac{|y_{ftn}|^2}{v_{ftn}} + \log v_{ftn}\right)$

- **Variance v constrained**
  - F-dimensional vector $\quad \mathbf{v}_{tn} = [v_{1tn}, \ldots, v_{Ftn}]^{\mathsf{T}}$



Variance update by similarity search



n-th separated signal

Variance update

$$\mathbf{v}_{tn} \leftarrow \gamma \breve{\mathbf{s}}_*$$

adjusted scale

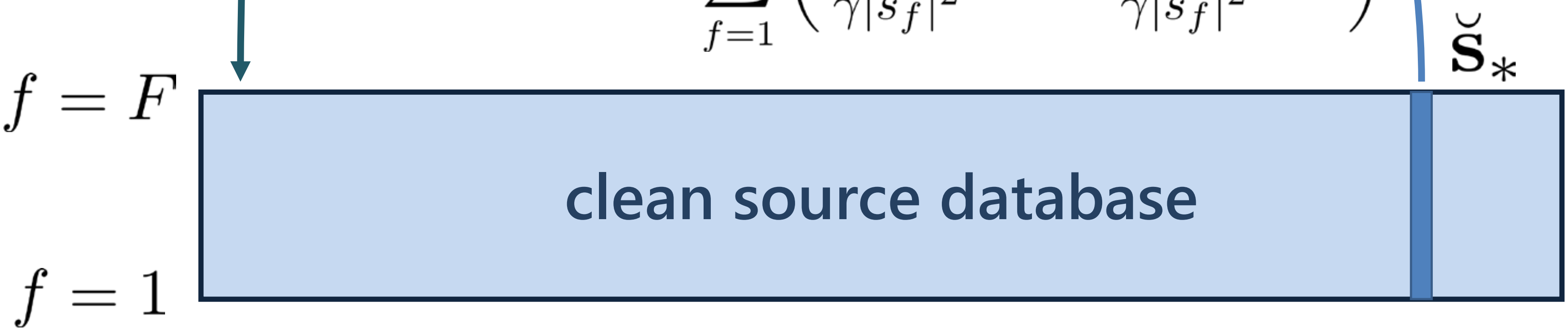
$$\gamma = \frac{1}{F} \sum_{f=1}^{F} \frac{|y_{ftn}|^2}{|s_f|^2}.$$

Query: power spectra vector

$$\breve{\mathbf{y}}_{tn} = [|y_{1tn}|^2, \ldots, |y_{Ftn}|^2]^\mathsf{T}$$

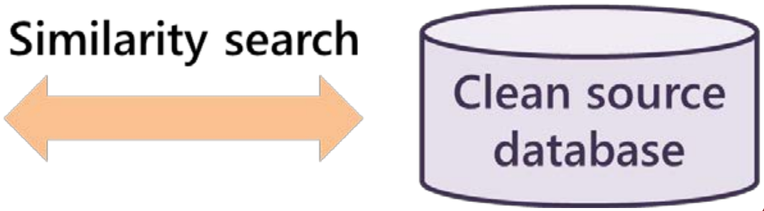
Measure: Itakura-Saito divergence

$$D_{IS}(\breve{\mathbf{y}}_{tn}, \gamma \breve{\mathbf{s}}) = \sum_{f=1}^{F} \left(\frac{|y_{ftn}|^2}{\gamma |s_f|^2} - \log \frac{|y_{ftn}|^2}{\gamma |s_f|^2} - 1 \right)$$



SSBSS Optimization

Variance update



$$\mathbf{y}_{ft} = \mathbf{W}_f \mathbf{x}_{ft}$$

Frequency-wise separation matrix update $\mathbf{W}_f$

Weighted covariance matrix

$$\mathbf{U}_{fn} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{v_{tn}}\mathbf{x}_{ft}\mathbf{x}_{ft}^{\mathsf{H}}$$

Solve HEAD: Hybrid Exact-Approximate Diagonalization

[Yeredor 2009]

$$\mathbf{w}_{fk}^{\mathsf{H}}\mathbf{U}_{fn}\mathbf{w}_{fn} = \delta_{kn}$$

N=2 case

$$\mathbf{w}_1^{\mathsf{H}}\mathbf{U}_1\mathbf{w}_1 = 1 \qquad \mathbf{w}_1^{\mathsf{H}}\mathbf{U}_2\mathbf{w}_2 = 0$$

$$\mathbf{w}_2^{\mathsf{H}}\mathbf{U}_1\mathbf{w}_1 = 0 \qquad \mathbf{w}_2^{\mathsf{H}}\mathbf{U}_2\mathbf{w}_2 = 1$$

[Ono 2011]

$$\mathbf{w}_{fn} \leftarrow (\mathbf{W}_f\mathbf{U}_{fn})^{-1}\mathbf{e}_n$$

$$\mathbf{w}_{fn} \leftarrow \frac{\mathbf{w}_{fn}}{\sqrt{\mathbf{w}_{fn}^{\mathsf{H}}\mathbf{U}_{fn}\mathbf{w}_{fn}}}$$

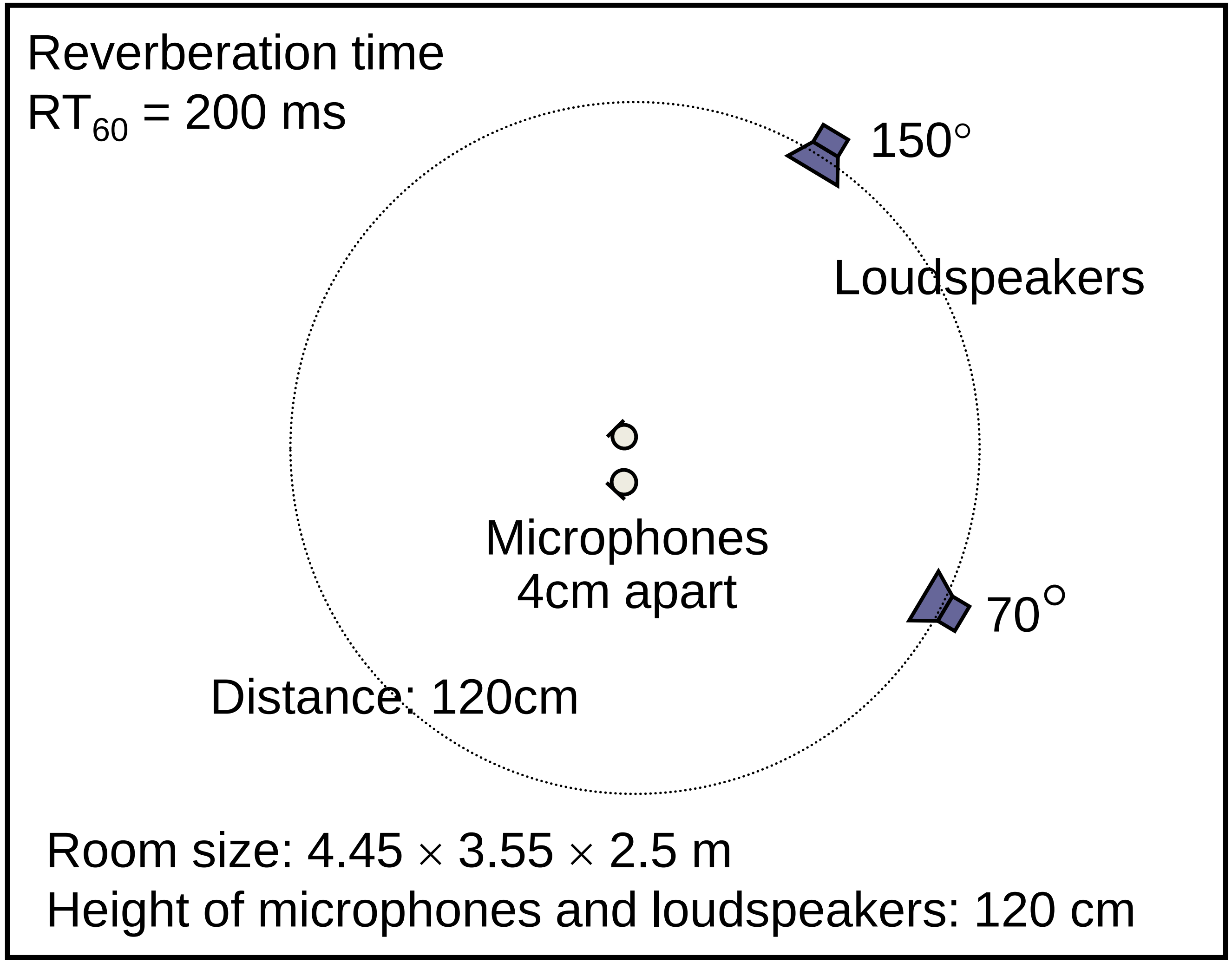
$$n = 1, \ldots, N$$

Outline

1. **Existing BSS methods**
   - Frequency-domain BSS
   - IVA, ICA, ILRMA
2. **Proposed method**
   - **SSBSS**: Similarity Search-based BSS
   - Differs in variance parameter updates
3. **Experiments**
   - Clean source databases: close and open
   - Convergence behavior
   - Computational time with a GPU

Experimental conditions

- **Sources:** 2-second speeches
- **Mixtures:** 32 cases
  Various combinations of 2-second speech signals

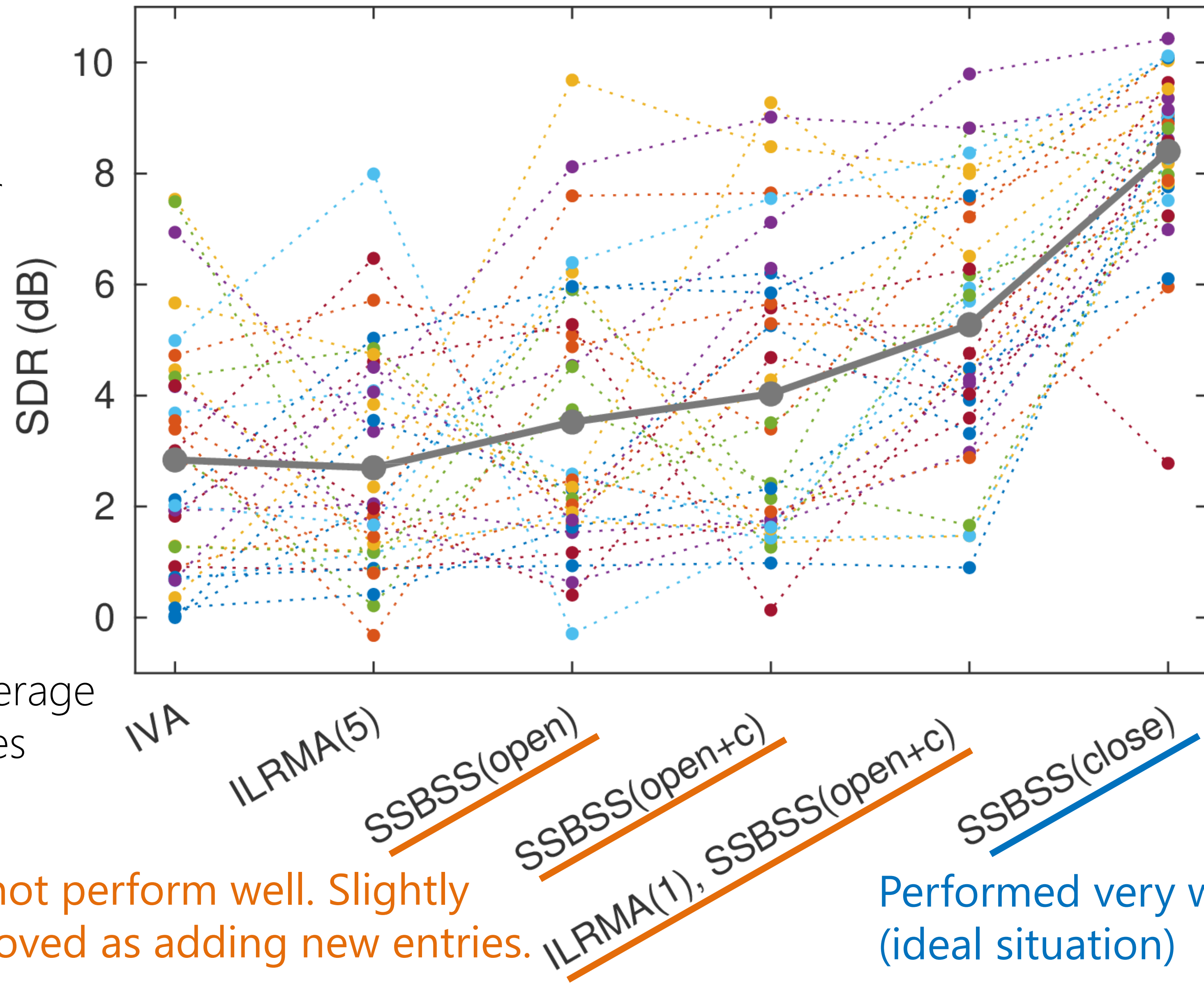


Clean source databases

- **close**
  - contained the sources used for mixtures
  - ideal situation for verifying the basic concept
- **open**
  - did not contain the source time frames used for mixtures
    - but contained the same speaker's different utterances
  - In some settings, new entries were added aiming for better performance

Separation performance

The higher the better



- Each dotted line corresponds to a mixture case
- Solid grey line represents the average of 32 mixture cases

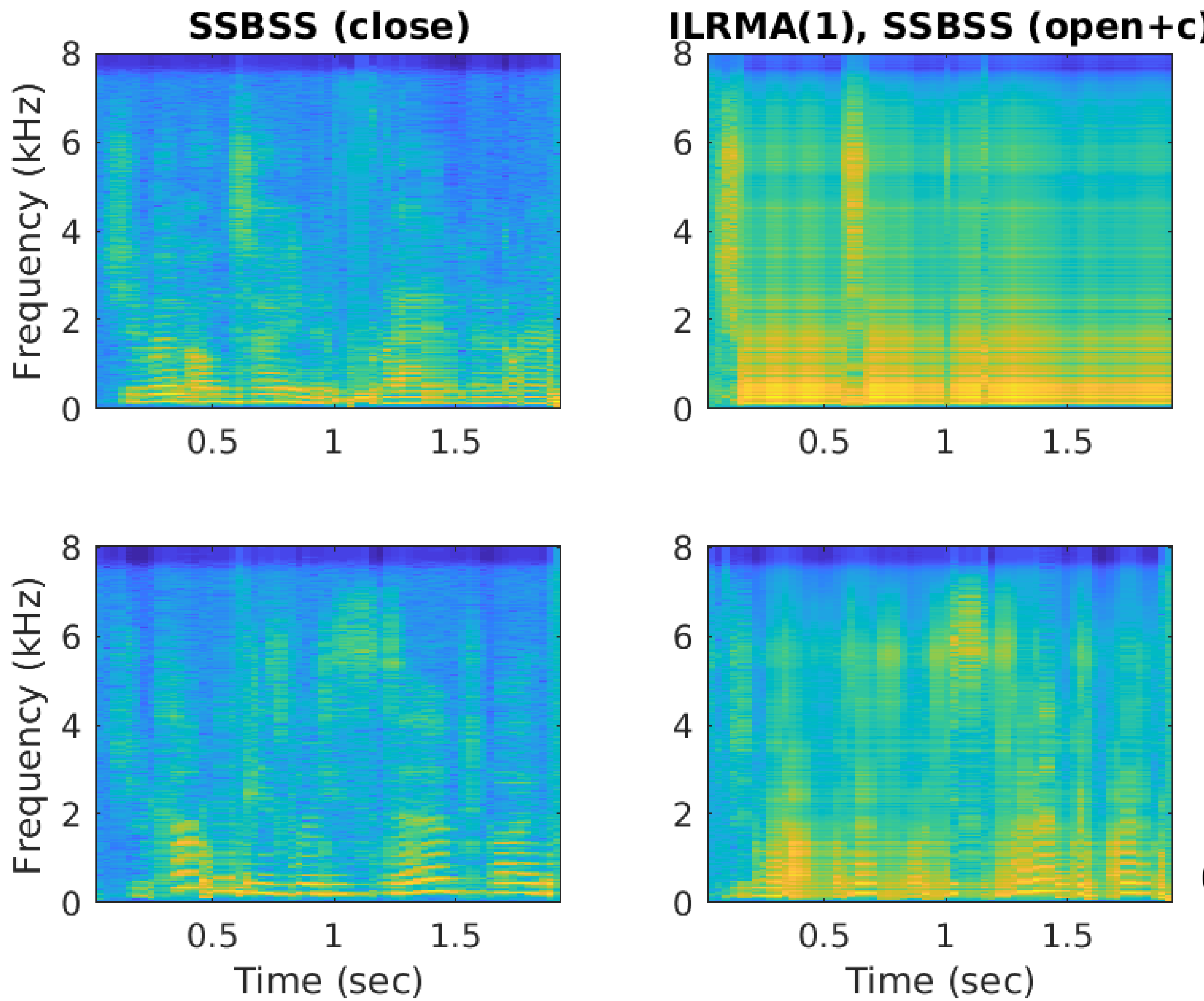
Did not perform well. Slightly improved as adding new entries.

Performed very well (ideal situation)

Variances & separated signals

Mixture

7.16 dB



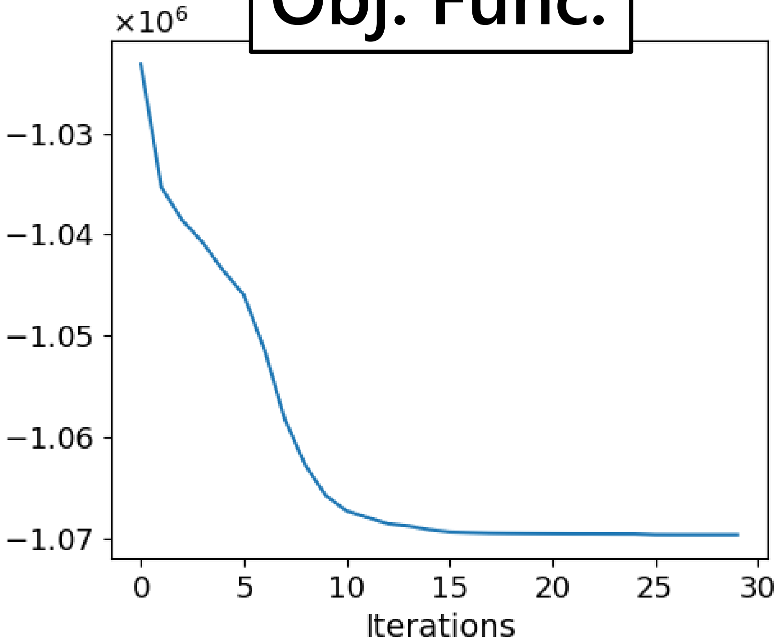
3.85 dB

9.79 dB

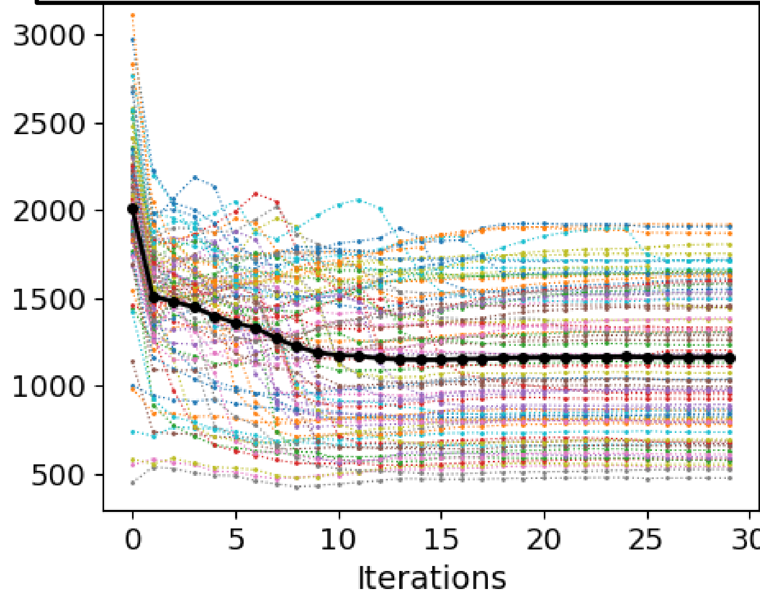
6.45 dB

Convergence & Computation

Obj. Func.



IS div. for time samples



30 iterations were sufficient

- **Execution time**
  - 20 seconds for 30 iterations and 2-second mixture
- **Similarity search executed on a GPU**
  - 158 queries (2 outputs × 79 time frames) for 30,000 entries with F=1025 dim. took around 230 ms.

Conclusion

- **Proposed SSBSS**
  - Searches clean database $\mathbb{S}$ for similar entries to $\check{\mathbf{y}}$
  - Updates variance parameters $\mathbf{v}$ with the result
- **Experimental results**
  - Short observation of 2 seconds
  - High performance with ideal close database
  - Open database lowered the performance
- **Future work**
  - Constructing better databases for open cases
  - Accelerating the search to handle larger databases