Paper #: 212

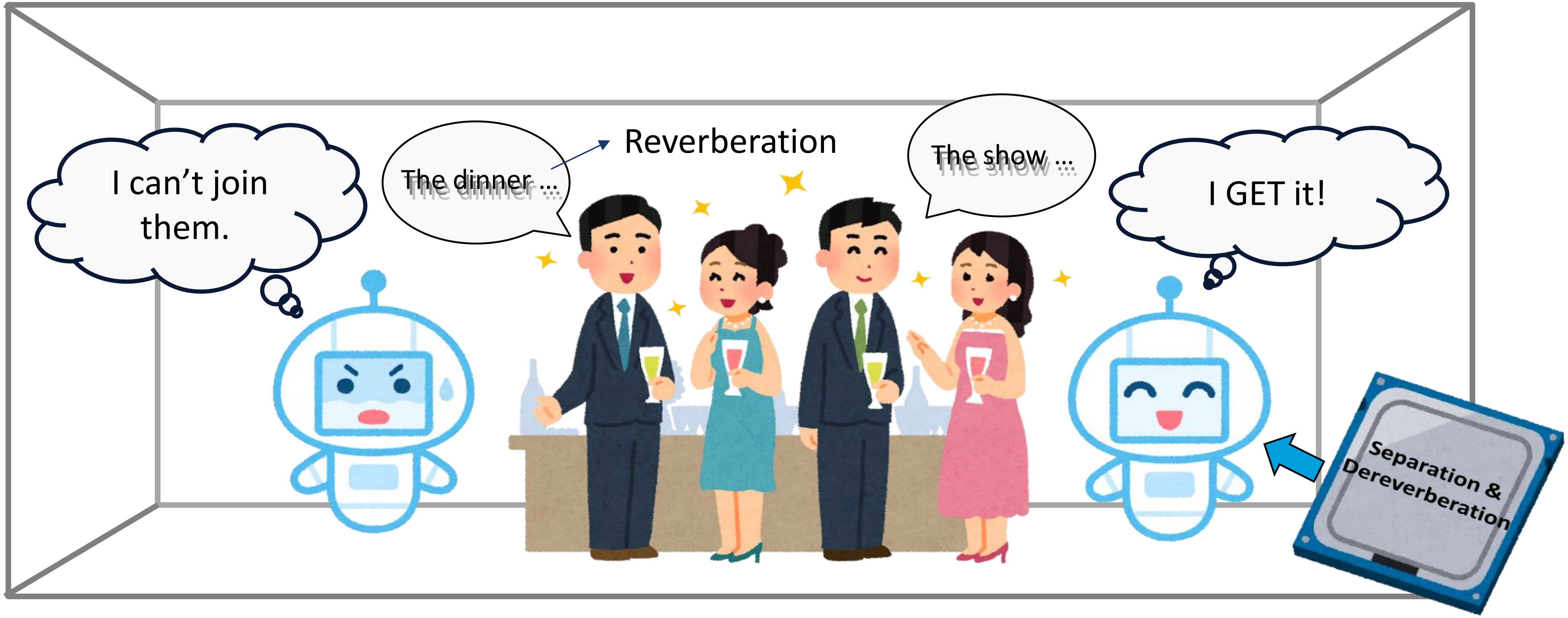
On Joint Dereverberation and Single Moving Source Separation with Online Source Steering

Yiting Zhang, Kaien Mo, Tetsuya Ueda, Yichen Yang, Shoji Makino

Waseda University, Japan

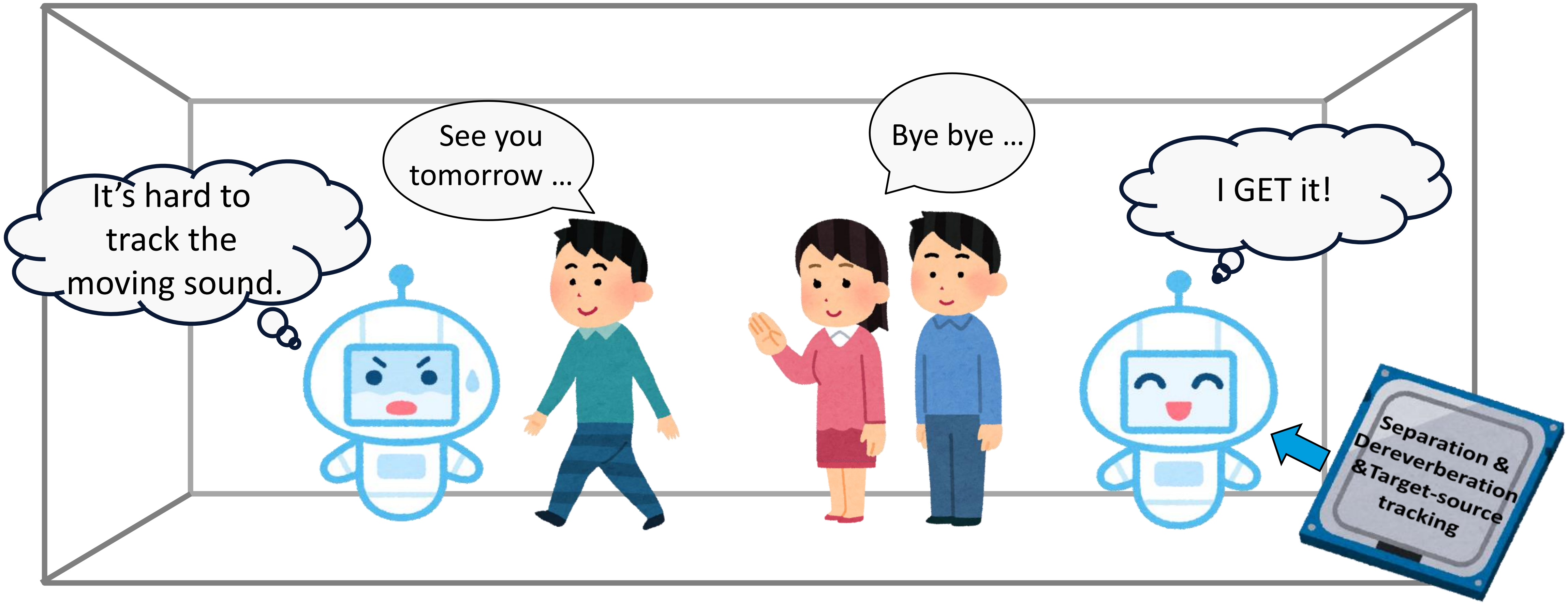
Research Background (1/2)

- The growing demand for speech processing and recognition
- Multi-speakers and reverberation significantly degrade their performance



Research Background (2/2)

- Separation performance becomes worse when a sound source is moving
- It is necessary to consider both reverberation and moving source



Problem formulation

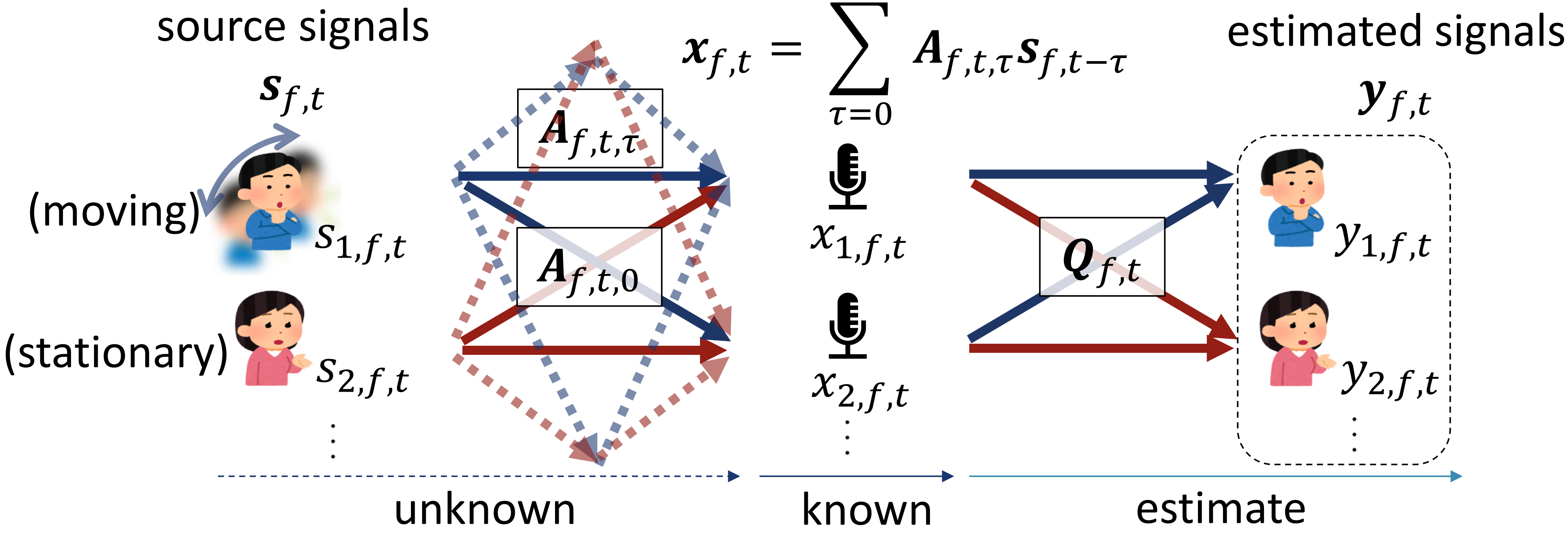
- Target-source tracking: only **one target source is moving** and other sources are stationary.

$\boldsymbol{A}_{f,t,\tau}$: mixing matrix
τ: time lag
L_a: # of mixing matrix

observed signals:

$$\boldsymbol{x}_{f,t} = \sum_{\tau=0}^{L_a-1} \boldsymbol{A}_{f,t,\tau}\boldsymbol{s}_{f,t-\tau}$$

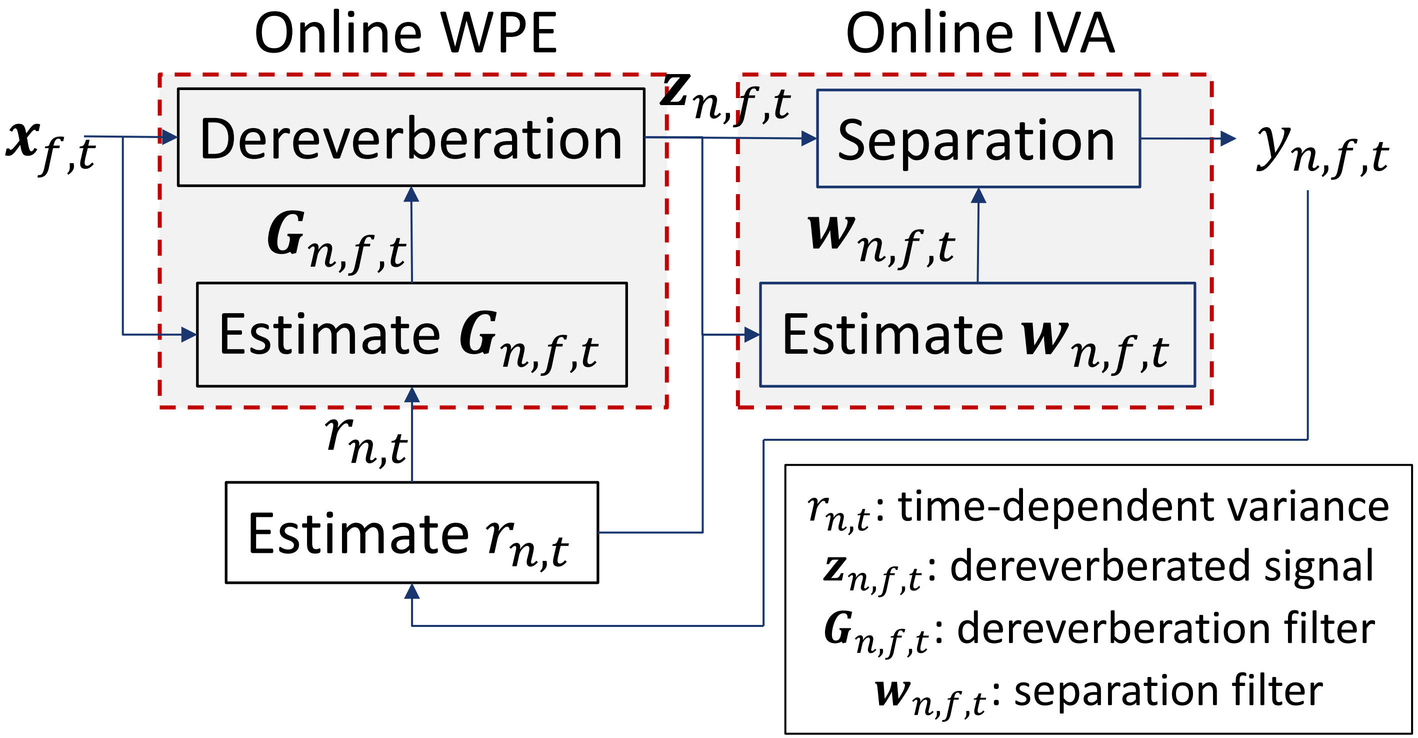
$\boldsymbol{Q}_{f,t}$: demixing matrix



t=1,…,T: time frames　f=1,…,F: frequency bins
n=1,…,N: # of source signals　m=1,…,M: # of microphones

Baseline Method: Online WPE-IVA

- Online Joint Optimization of Weighted Prediction Error and Independent Vector Analysis (Online WPE-IVA) [Ueda+ 2021; Mo+ 2024]
  - Convolutional Beamformer (CBF) with source-wise factorization



$$\boldsymbol{y}_{f,t} = \boldsymbol{Q}_{f,t,0}^{\mathrm{H}} \boldsymbol{x}_{f,t} + \sum_{\tau=D}^{D+L-1} \boldsymbol{Q}_{f,t,\tau}^{\mathrm{H}} \boldsymbol{x}_{f,t-\tau}$$

$$\boldsymbol{z}_{n,f,t} = \boldsymbol{x}_{f,t} - \boldsymbol{G}_{n,f,t}^{\mathrm{H}} \overline{\boldsymbol{x}}_{f,t}$$

$$y_{n,f,t} = \boldsymbol{w}_{n,f,t}^{\mathrm{H}} \boldsymbol{z}_{n,f,t}$$

D: prediction delay

L: number of demixing matrix

$\overline{\boldsymbol{x}}_{f,t} = \left[\boldsymbol{x}_{f,t-D}^{\mathrm{T}}, \dots, \boldsymbol{x}_{f,t-D-L+1}^{\mathrm{T}}\right]^{\mathrm{T}}$

  - Online: filters are updated after every new frame

Baseline Method: Online WPE-IVA

- Update rule of Online WPE-IVA

$\boldsymbol{k}_{n,f,t}$: Kalman gain obtained by past observed signals

Online WPE part: $\boldsymbol{G}_{n,f,t} = \boldsymbol{G}_{n,f,t-1} + \boldsymbol{k}_{n,f,t}\boldsymbol{z}^{\mathrm{H}}_{n,f,t}$

α: forgetting vector

Online IVA part: **Matrix computation!** $\boldsymbol{V}_{n,f,t} = \alpha \boldsymbol{V}_{n,f,t-1} + (1-\alpha)\frac{\boldsymbol{z}_{n,f,t}\boldsymbol{z}^{\mathrm{H}}_{n,f,t}}{r_{n,t}}$

①using Iterative Projection (IP) [Ueda+ 2021]

$$\boldsymbol{w}_{n,f,t} = (\boldsymbol{W}_{f,t-1}\boldsymbol{V}_{n,f,t})^{-1}\boldsymbol{e}_{n,f}$$

$$\boldsymbol{w}_{n,f,t} = (\boldsymbol{w}^{\mathrm{H}}_{n,f,t}\boldsymbol{V}_{n,f,t}\boldsymbol{w}_{n,f,t})^{-1/2}\boldsymbol{w}_{n,f,t}$$

$$(\boldsymbol{W}_{f,t} = [\boldsymbol{w}_{1,f,t}, \ldots, \boldsymbol{w}_{N,f,t}]^{\mathrm{H}})$$

$\boldsymbol{W}_{f,t}$: separation matrix for $n = 1 \ldots N$

one-hot update

②using Iterative Source Steering (ISS) [Mo+ 2024]

$$u_{n,f,t} = \boldsymbol{w}^{\mathrm{H}}_{i,f,t-1}\boldsymbol{V}_{n,f,t}\boldsymbol{w}_{n,f,t-1}$$

$$d_{n,f,t} = \boldsymbol{w}^{\mathrm{H}}_{i,f,t-1}\boldsymbol{V}_{n,f,t}\boldsymbol{w}_{i,f,t-1}$$

$$\boldsymbol{w}_{n,f,t} = \begin{cases} \boldsymbol{w}_{n,f,t-1} - \dfrac{u_{n,f,t}}{d_{n,f,t}}\boldsymbol{w}_{i,f,t-1}, n \neq i \\ d_{i,f,t}^{-\frac{1}{2}}\boldsymbol{w}_{i,f,t-1}, n = i \end{cases}$$

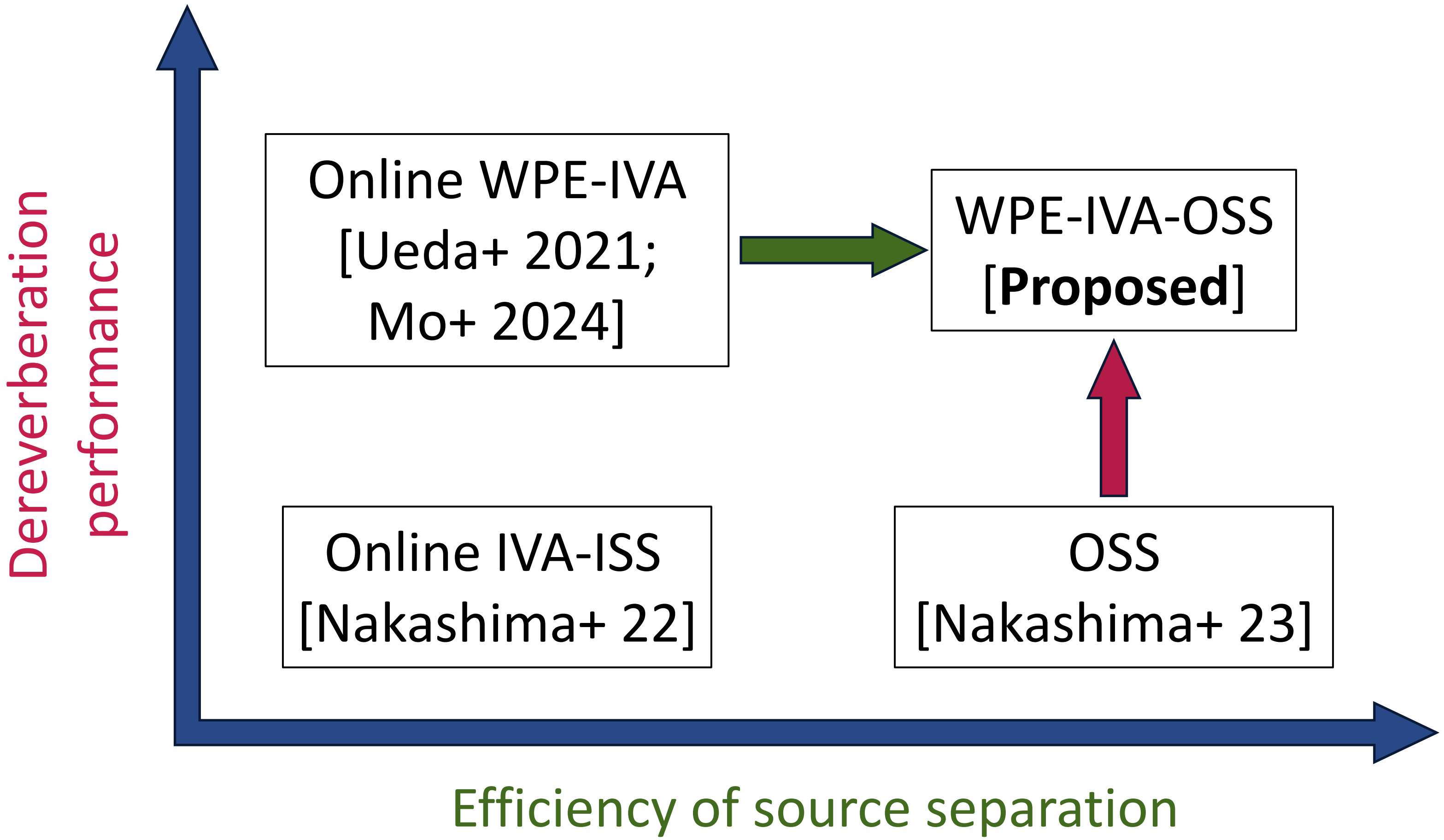
rank-1 update

i: moving source

Motivation and Contribution

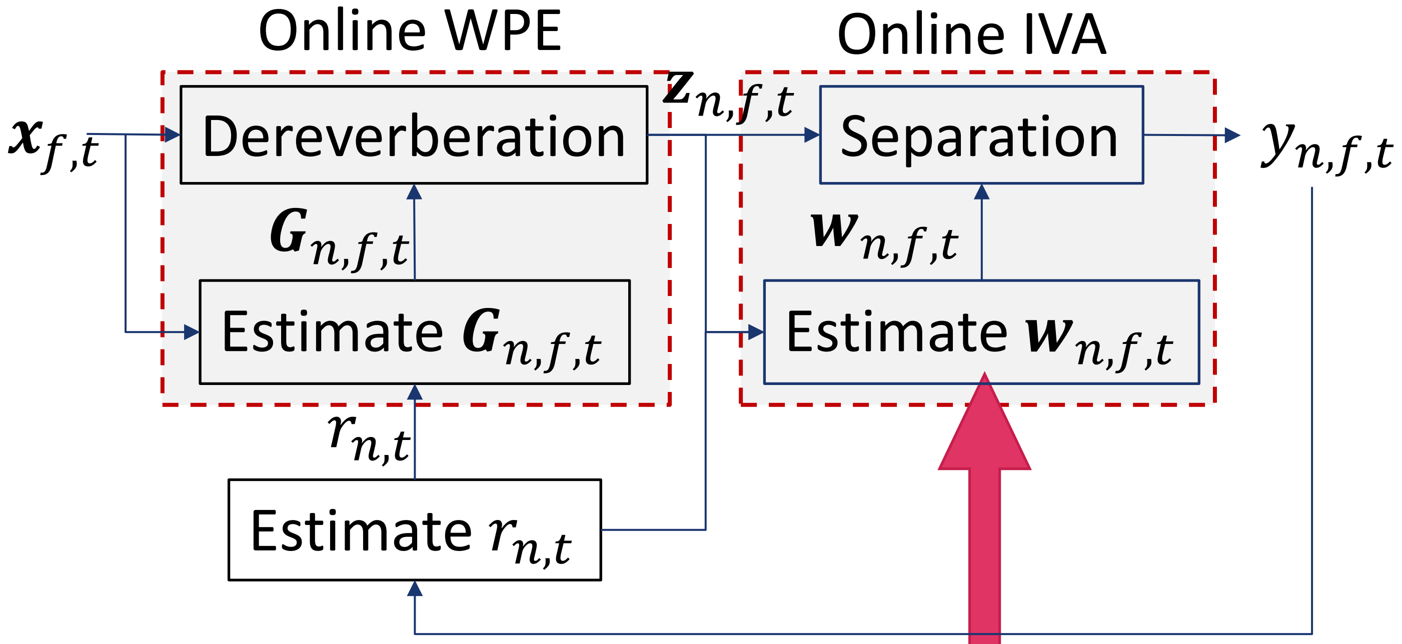
- Baseline Method: Online WPE-IVA [Ueda+ 2021; Mo+ 2024]
  - Advantage:
    - Suitable for highly reverberant environments
    - Suitable for real-time processing under time-variant environments
  - Shortcoming:
    - Relative slow update algorithm in separation filter
- Contribution
  1. Promote computational efficiency by apply Online Source Steering (OSS, explained later) [Nakashima+ 23].
  2. Apply Geometric Constraint (GC) [Parra+ 03] to fix permutation before using our proposed method.

Motivation and Contribution



Proposed Method: WPE-IVA-OSS

- Using Online Source Steering (OSS) algorithm [Nakashima+ 2023] to replace IP or ISS



  – Advantage:
    - Relative fast update algorithm in separation filter

OSS algorithm:
matrix computation is avoided

Proposed Method: WPE-IVA-OSS

- Update rule of WPE-IVA-OSS

Online WPE part: $\boldsymbol{G}_{n,f,t} = \boldsymbol{G}_{n,f,t-1} + \boldsymbol{k}_{n,f,t}\boldsymbol{z}_{n,f,t}^{\mathrm{H}}$

Online IVA part: **No matrix computation!** $\boldsymbol{V}_{n,f,t} = \alpha \boldsymbol{V}_{n,f,t-1} + (1-\alpha)\frac{\boldsymbol{z}_{n,f,t}\boldsymbol{z}_{n,f,t}^{\mathrm{H}}}{r_{n,t}}$

Coefficients for updating $\boldsymbol{w}_{n,f,t-1}$:

$$u_{n,f,t} = \boldsymbol{w}_{i,f,t-1}^{\mathrm{H}} \boldsymbol{V}_{n,f,t} \boldsymbol{w}_{n,f,t-1}$$

$$d_{n,f,t} = \boldsymbol{w}_{i,f,t-1}^{\mathrm{H}} \boldsymbol{V}_{n,f,t} \boldsymbol{w}_{i,f,t-1}$$

using **O**nline **S**ource **S**teering(OSS) [Nakashima+ 2023]

$$u_{n,f,t} = \alpha \frac{u_{n,f,t-1}}{u_{i,f,t-1}} + \frac{(1-\alpha)\left|\hat{y}_{i,f,t}\right|^2}{r_{n,t}}, d_{n,f,t} = \frac{(1-\alpha)\hat{y}_{i,f,t}y_{n,f,t}^{*}}{r_{n,t}}$$

$u_{n,f,t}$, $d_{n,f,t}$ are recursively updated without $\boldsymbol{V}_{n,f,t}$

Experiment

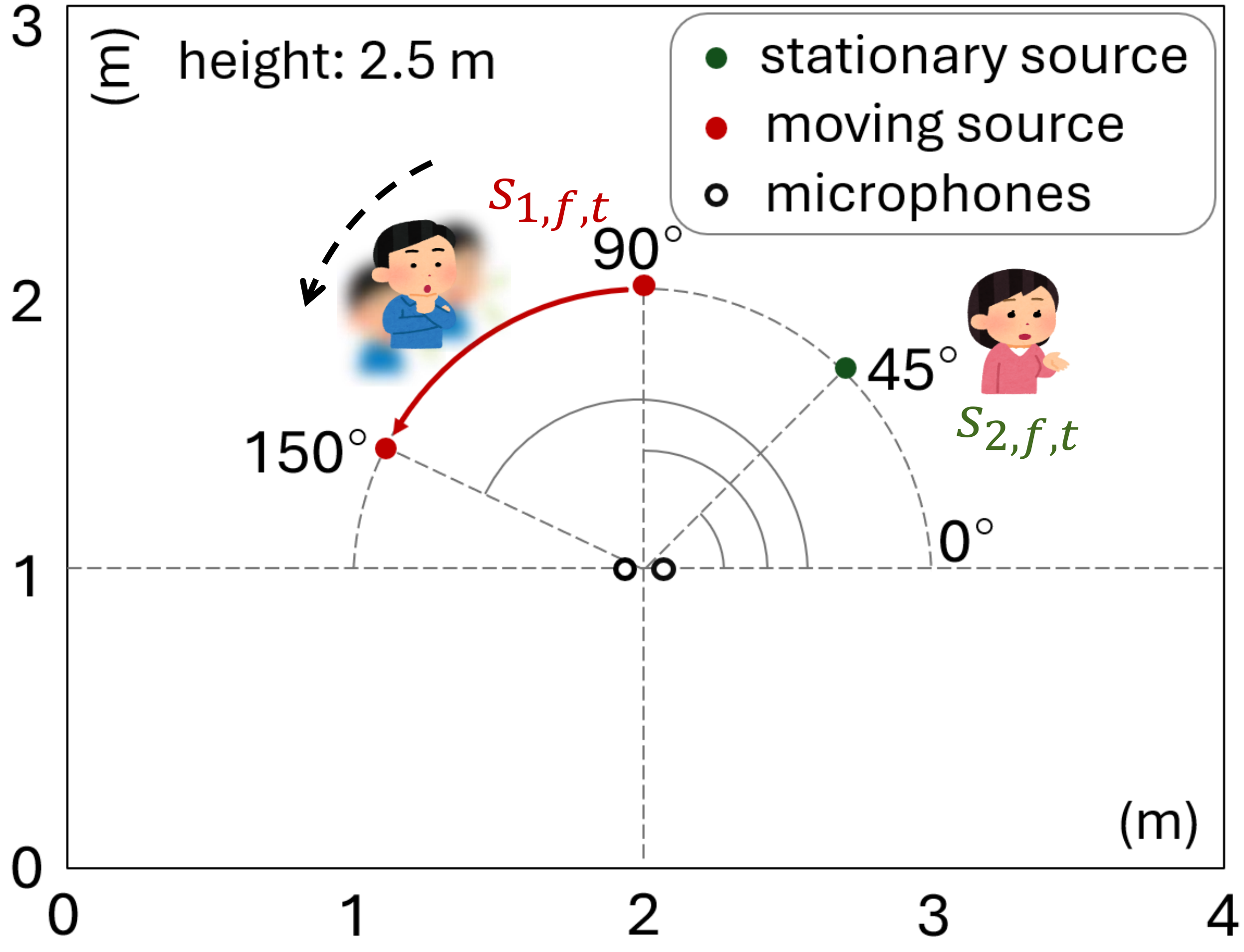
- Motivation: evaluate our proposed method's
  1. Separation performance
  2. Computational efficiency

- Evaluation Criteria
  - ΔSDR [dB]: Improvement of source-to-distortions ratio (SDR_{output} - $SDR_{observed}$)
  - Runtime [s]

- Compared Methods
  - Online WPE-IVA-IP [Ueda+ 2021]
  - Online WPE-IVA-ISS [Mo+ 2024]
  - WPE-IVA-OSS [**Proposed**]

Experiment

- Settings

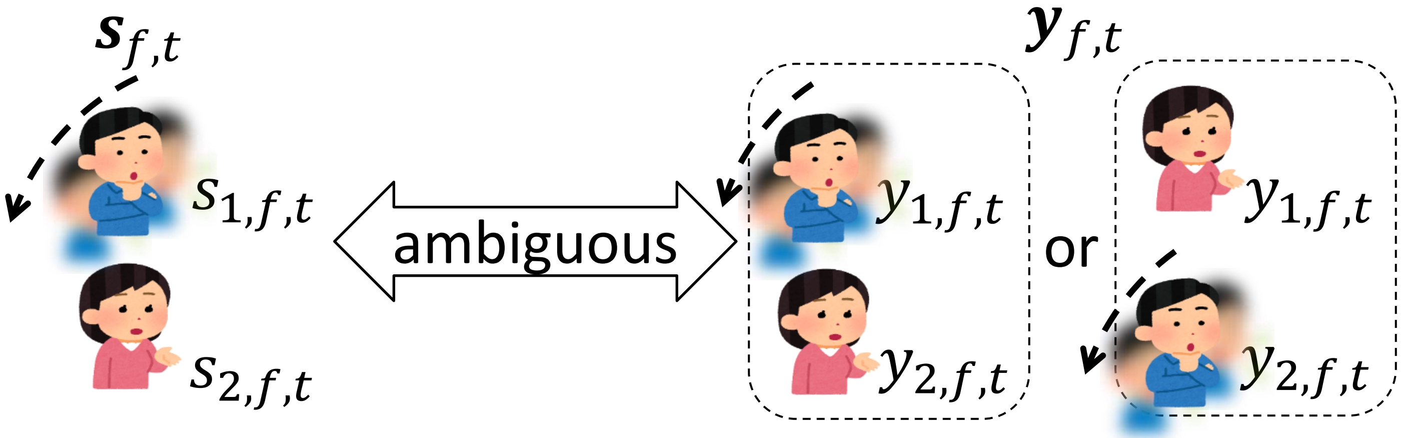


Mixing conditions	
Speech datasets	ATR503 database set B
Reverberation time (RT60)	600 ms
Number of microphone (= Number of source)	2
Number of dataset pair	20 pairs

Experimental conditions	
Sampling frequency	16 kHz
STFT window/shift length	1024/256
STFT window function	Hann
α (forgetting factor)	0.99
β (forgetting factor)	0.96
L (filter order)	10
D (prediction delay)	2
Initial values of $\boldsymbol{G}_{n,f,t}$	Zero matrix
Initial values of $\boldsymbol{W}_{f,t}$	Identity matrix

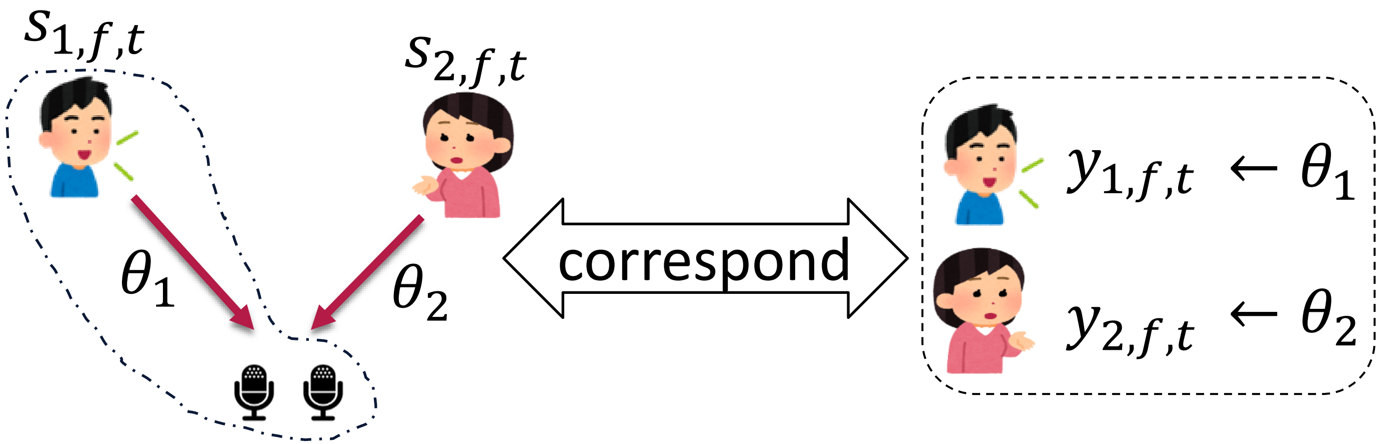
Experiment

- Online WPE-IVA with Geometric Constraint (Online WPE-GCIVA) [Mo+ 2023]
  - Permutation problem of estimated signals



Random output permutation

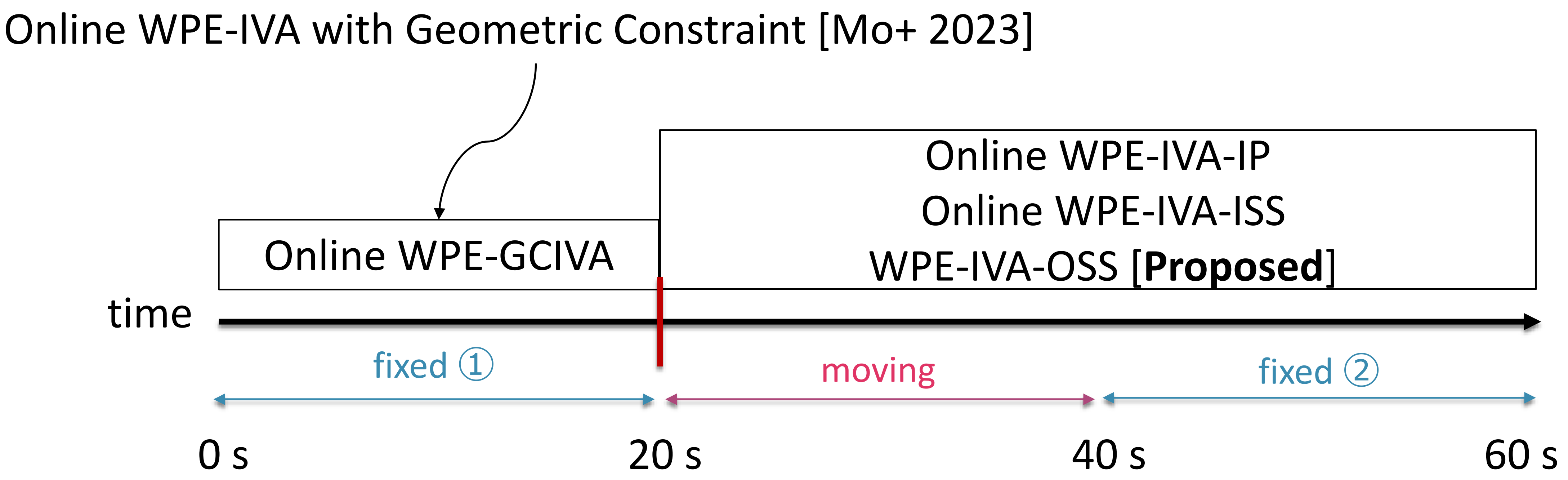
  - Geometric Constraint (GC): Guide separation filters to suppress signal in direction θ



Fixed output permutation

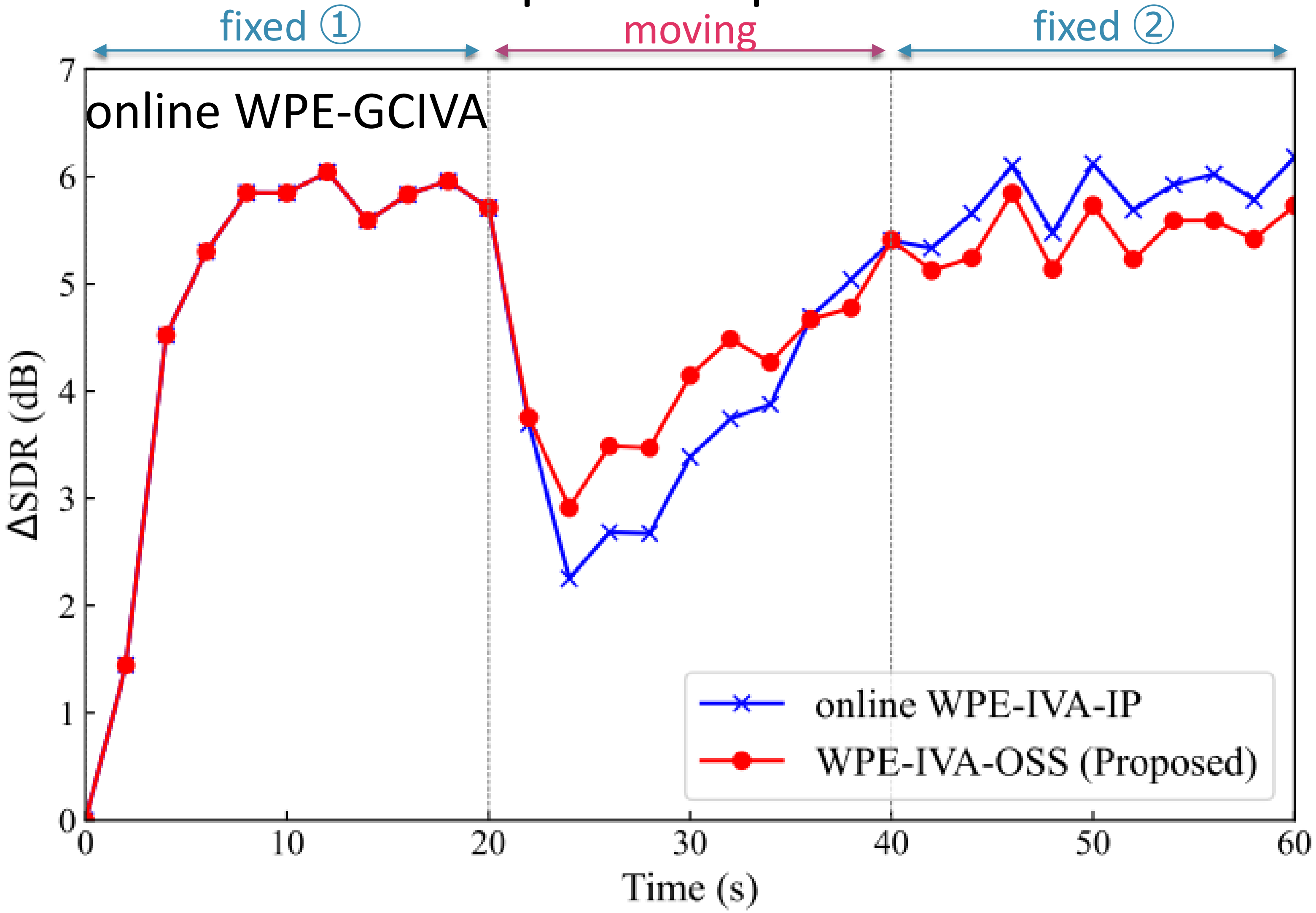
Experiment

- Settings



Experiment

- Result 1: Source separation performance



- WPE-IVA-OSS converged faster than online WPE-IVA-IP.
- After the convergence, WPE-IVA-OSS performed slightly worse than online WPE-IVA-IP.

Experiment

- Result 2: Runtime

Online WPE-IVA-IP	Online WPE-IVA-ISS	WPE-IVA-OSS
2.4124 s	1.2775 s	0.4808 s

  - Only include last 40s because updates of first 20s are the same
  - Only include separation part because other parts are the same

Conclusions

- We integrated OSS update algorithm to replace conventional IP- and ISS-based one in online WPE-IVA.
- Our proposed method fully took advantages of online WPE-IVA.
- Our proposed method has low computational cost and no matrix computation in separation filter.
- Our proposed method is computationally more efficient while achieving the same or at least comparable separation performance.

Thank you for your listening!