VECTORWISE COORDINATE DESCENT ALGORITHM FOR SPATIALLY REGULARIZED INDEPENDENT LOW-RANK MATRIX ANALYSIS

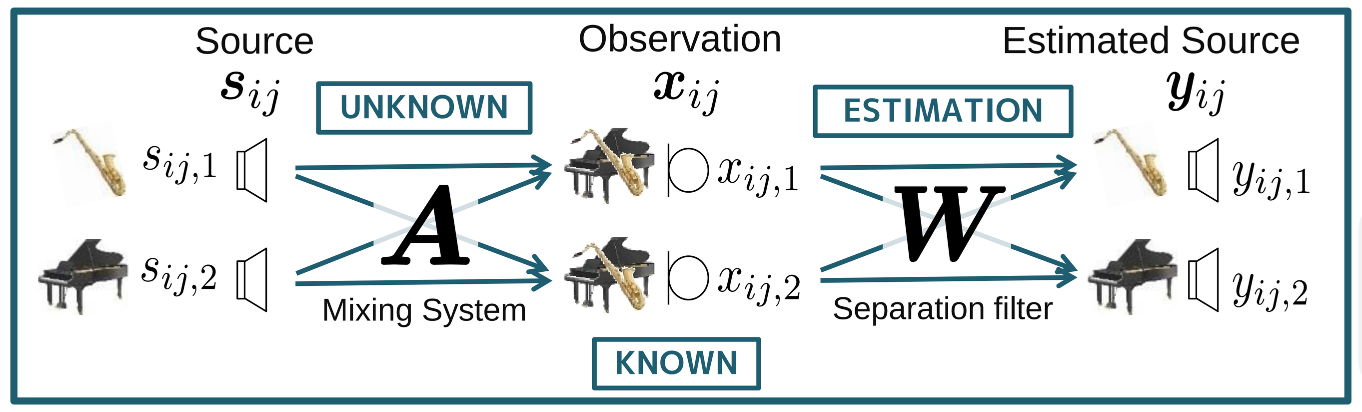
Yoshiki Mitsui [1], Norihiro Takamune [1], Daichi Kitamura [1], Hiroshi Saruwatari [1], Yu Takahashi [2], Kazunobu Kondo [2]

[1] The University of Tokyo, Tokyo, Japan [2] YAMAHA Corporation, Shizuoka, Japan

1. SUMMARY

Independent low-rank matrix analysis (ILRMA)

- **A novel blind audio source separation method** [Kitamura+ 2016]



i : integral index of frequency
j : integral index of time

Problem: block permutation problem (misalignment in the low- and high-frequency bands) especially in speech/speech separation

Proposal: spatial-model-based regularization method for ILRMA with null beamformer and new optimization algorithm (vectorwise coordinate descent: VCD)

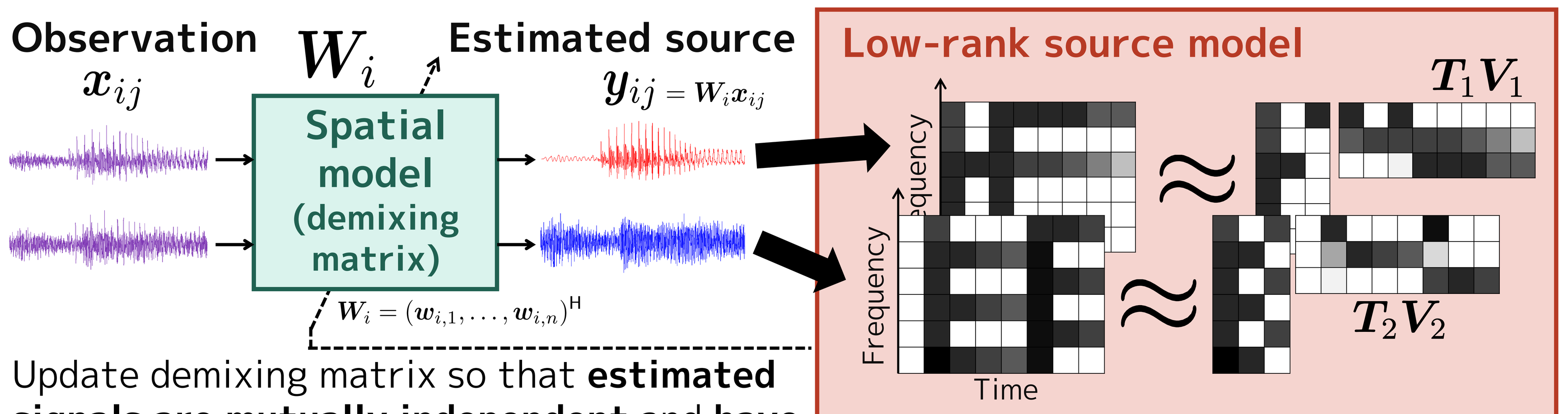
Result: proposed approach can improve the separation accuracy and stability in speech/speech separation.

2. ILRMA

$$D_{i,n} = \frac{1}{J}\sum_j \frac{x_{ij}x_{ij}^{\mathsf{H}}}{\sum_l t_{il,n}v_{lj,n}}$$

$$\mathcal{J} = -\sum_i \log|\det W_i|^2 + \sum_{i,n}\left[w_{i,n}^{\mathsf{H}} D_{i,n} w_{i,n} + \log \sum_l t_{il,n}v_{lj,n}\right]$$

Spatial model (IVA) Low-rank source model (NMF)



Update demixing matrix so that **estimated signals are mutually independent** and have **low-rank structure** in time-frequency domain

l : integral index of NMF basis

Conventional update rule for demixing filters $w_{i,n}$: iterative projection (IP) [Ono 2011]

$$u_{i,n} = D_{i,n}^{-1} W_i^{-1} e_n, \quad w_{i,n} \leftarrow \frac{u_{i,n}}{\sqrt{u_{i,n}^{\mathsf{H}} D_{i,n} u_{i,n}}}$$

e_n : unit vector (nth element is unity)

- IP can be applied to only $w_{i,n}^{\mathsf{H}} D_{i,n} w_{i,n} - \log|\det W_i|$

3. PROPOSED METHOD

Introduce the beamformer-based regularizer into ILRMA

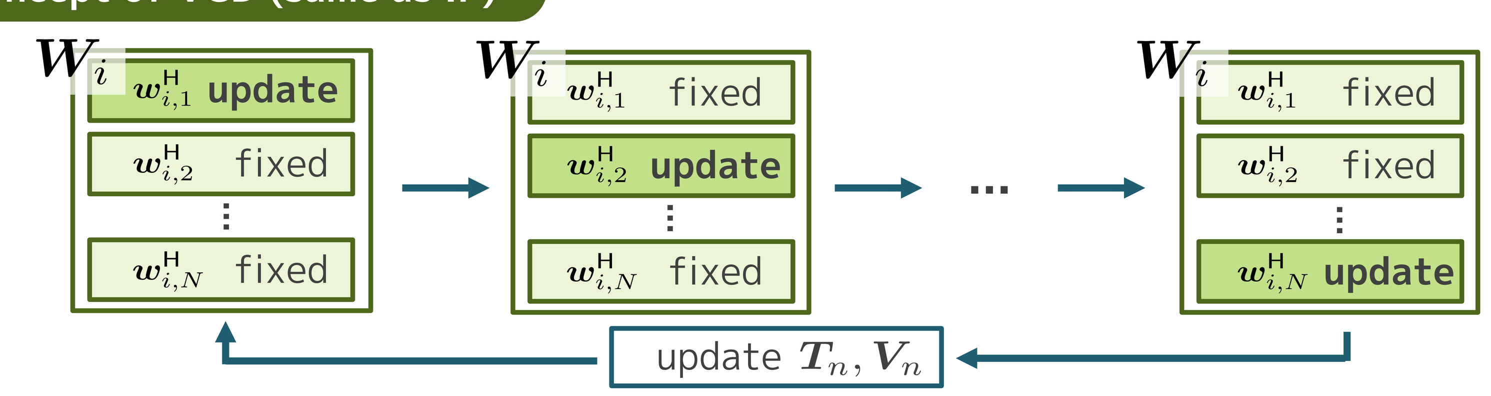
$$\mathcal{J}_{\mathrm{R}} = \mathcal{J} + \sum_{i,n}\lambda_n \|w_{i,n} - \widehat{w}_{i,n}\|^2$$

$\widehat{w}_{i,n}$: beamformer-based supervisor
$\widehat{D}_{i,n} = D_{i,n} + \lambda_n I_N$

$$= \sum_i\left[\sum_n\left(w_{i,n}^{\mathsf{H}}\widehat{D}_{i,n}w_{i,n} - \lambda_n\widehat{w}_{i,n}^{\mathsf{H}}w_{i,n} - \lambda_n w_{i,n}^{\mathsf{H}}\widehat{w}_{i,n}\right) - \log|\det W_i|^2\right] + \mathcal{C}$$

- Due to the linear terms $w_{i,n}^{\mathsf{H}}\widehat{w}_{i,n}$ → **cannot apply IP!**
- By introducing **cofactor expansion of W_i**, **we can optimize $\mathcal{J}_{\mathrm{R}}$ with VCD** (IP-like optimization method)

Concept of VCD (same as IP)



Derivation of VCD

- $B_i = (b_{i,1}, \ldots, b_{i,N})$: adjugate matrix of W_i
- Since $b_{i,n}$ is independent of $w_{i,n}$,

$$\frac{\partial \log|\det W_i|^2}{\partial w_{i,n}^*} = \frac{\partial \log|w_{i,n}^{\mathsf{H}}b_{i,n}|^2}{\partial w_{i,n}^*} = \frac{b_{i,n}}{w_{i,n}^{\mathsf{H}}b_{i,n}}$$

- By replacing $\beta_{i,n} = 1/(w_{i,n}^{\mathsf{H}}b_{i,n})$, the equation $\partial\mathcal{J}_{\mathrm{R}}/\partial w_{i,n}^* = \mathbf{0}$ can be transformed as a following quadratic equation w.r.t. $\beta_{i,n}$:

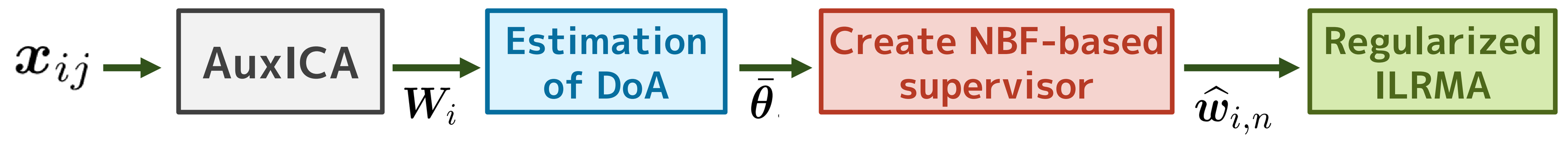
$$b_{i,n}^{\mathsf{H}}\widehat{D}_{i,n}^{-1}b_{i,n}|\beta_{i,n}|^2 + \lambda_n\widehat{w}_{i,n}^{\mathsf{H}}\widehat{D}_{i,n}^{-1}b_{i,n}\beta_{i,n} - 1 = 0$$

- Finally, we obtain the following update rules of $w_{i,n}$:

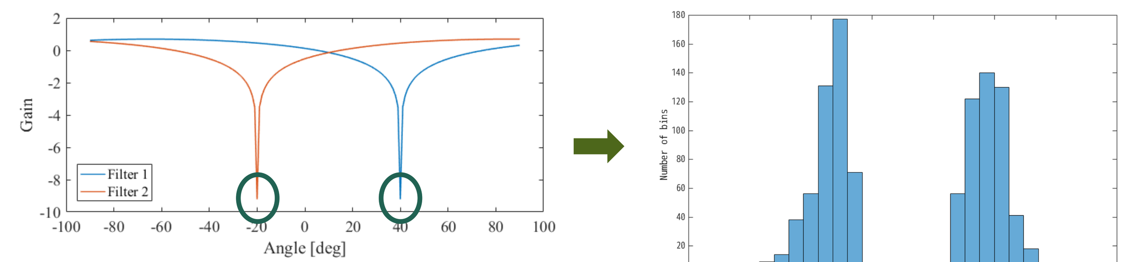
$$u_{i,n} = \widehat{D}_{i,n}^{-1}W_i^{-1}e_n,$$
$$\widehat{u}_{i,n} = \lambda_n\widehat{D}_{i,n}^{-1}\widehat{w}_{i,n},$$
$$r_{i,n} = u_{i,n}^{\mathsf{H}}\widehat{D}_{i,n}u_{i,n},$$
$$\widehat{r}_{i,n} = u_{i,n}^{\mathsf{H}}\widehat{D}_{i,n}\widehat{u}_{i,n},$$
$$w_{i,n} \leftarrow \begin{cases} \dfrac{u_{i,n}}{\sqrt{r_{i,n}}} + \widehat{u}_{i,n} & (\text{if } \widehat{r}_{i,n} = 0) \\ \dfrac{\widehat{r}_{i,n}}{2r_{i,n}}\left(-1 + \sqrt{1 + \dfrac{4r_{i,n}}{|\widehat{r}_{i,n}|^2}}\right)u_{i,n} + \widehat{u}_{i,n} & (\text{otherwise}) \end{cases}$$

Supervisor $\widehat{w}_{i,n}$ based on null beamformer (NBF)

- From previous studies, **NBF-based supervisor is effective.**
- We use AuxICA [Ono+ 2010] and k-means clustering techniques to estimate the direction of sources.



- How to estimate $\bar{\theta}$ from W_i:



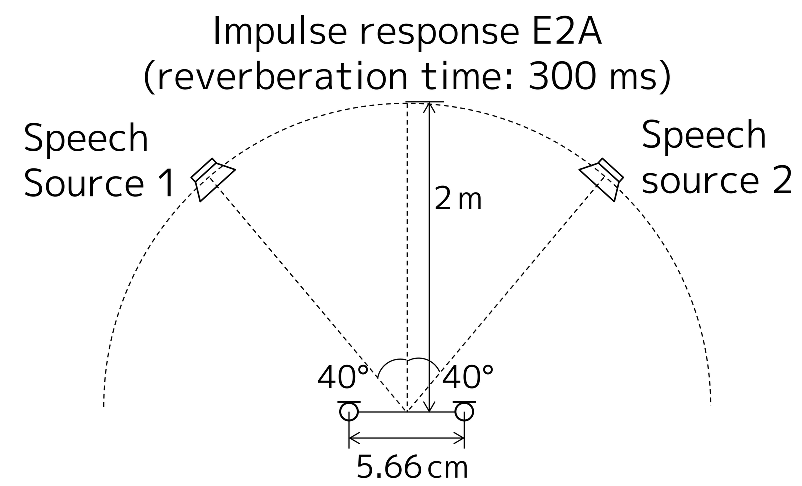
Calculate the null direction from $w_{i,n}$
※ the null direction is different in each frequency bin

Make a histogram and k-means clustering

Estimated DoA: (-34.61°, 18.07°)

4. EXPERIMENT

Experimental conditions



Recording environment

Initial value of W	(1) Identity matrix (2)(3) NBF
Initial value of T, V	Unit random number
Spatial regularization	(1)(2) Off (3) On
Number of NMF bases	1,2,3,5,10,15,20,25,30
Evaluation criteria	Avg. improvement of signal-to-distortion ratio (SDRi) (20-time trial)
Speech sources	Obtained from SiSEC UND datasets (4 pairs)

Result: Average SDR improvement [dB]

	Regularization weight	Number of NMF bases								
		1	2	3	5	10	15	20	25	30
(1)	-	6.07	7.35	6.23	6.26	5.60	5.45	5.14	4.97	5.06
(2)	0	8.04	9.38	10.28	10.95	11.34	11.60	11.67	11.66	11.51
(3) Prop.	0.3	**8.06**	9.76	10.95	12.04	12.36	12.55	**12.40**	12.34	12.25
	1	**8.06**	10.13	11.23	**12.16**	**12.40**	**12.62**	12.35	**12.43**	**12.27**
	3	**8.06**	**10.43**	**11.40**	12.07	12.22	12.40	12.23	12.20	12.04

Comparison of normalized calculation time per an iteration

AuxIVA [Ono 2011]	MNMF [Sawada+ 2013]	t-MNMF [K. Kitamura+ 2016]	ILRMA [D. Kitamura+ 2016]	Proposed ILRMA
0.69	61.99	71.95	1.00	**1.61**

Spatially regularized ILRMA can improve the speech/speech separation quality efficiently!