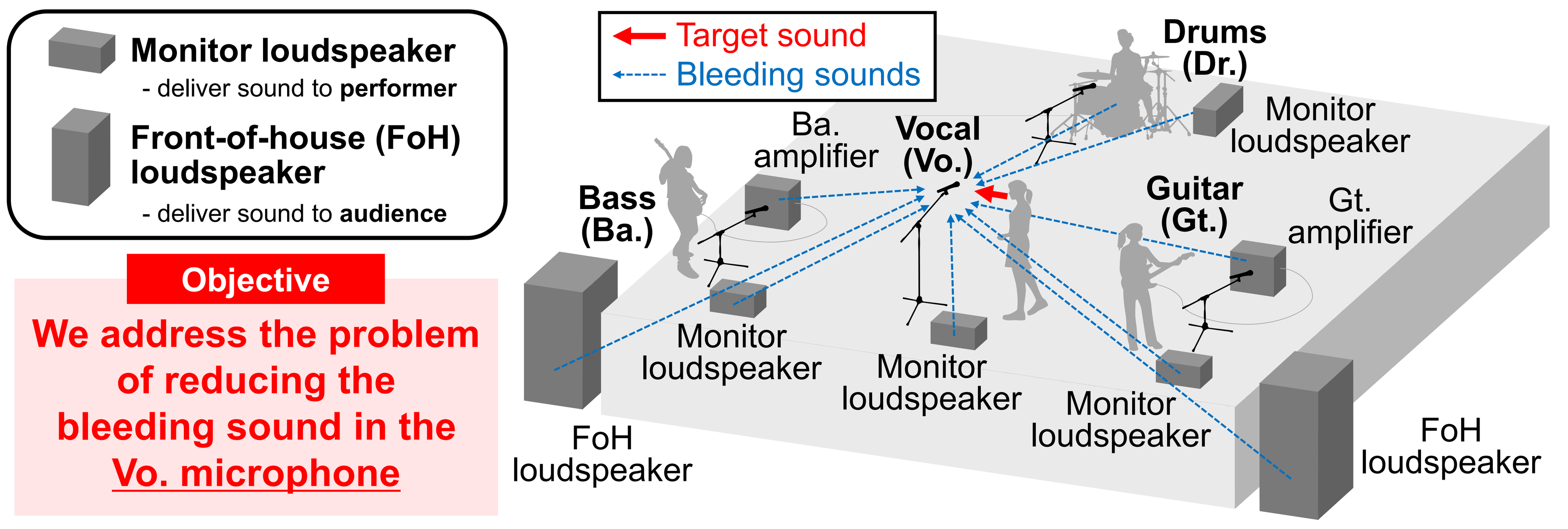
228

Demixing Filter Estimation for Bleeding-Sound Reduction of a Vocal Microphone

Soushi Taninomiya, Daichi Kitamura (NIT Kagawa, Japan), Norihoro Takamune, Kouei Yamaoka, Hiroshi Saruwatari (UTokyo, Japan),Yu Takahashi, Kazunobu Kondo, Hayato Yamakawa (Yamaha, Japan)

1. Research Background

- **Bleeding sound**
  - In live music performances, it's necessary to clearly capture **Vocal source**
  - But **non-target (guitar, bass, and drums) sources** are also captured

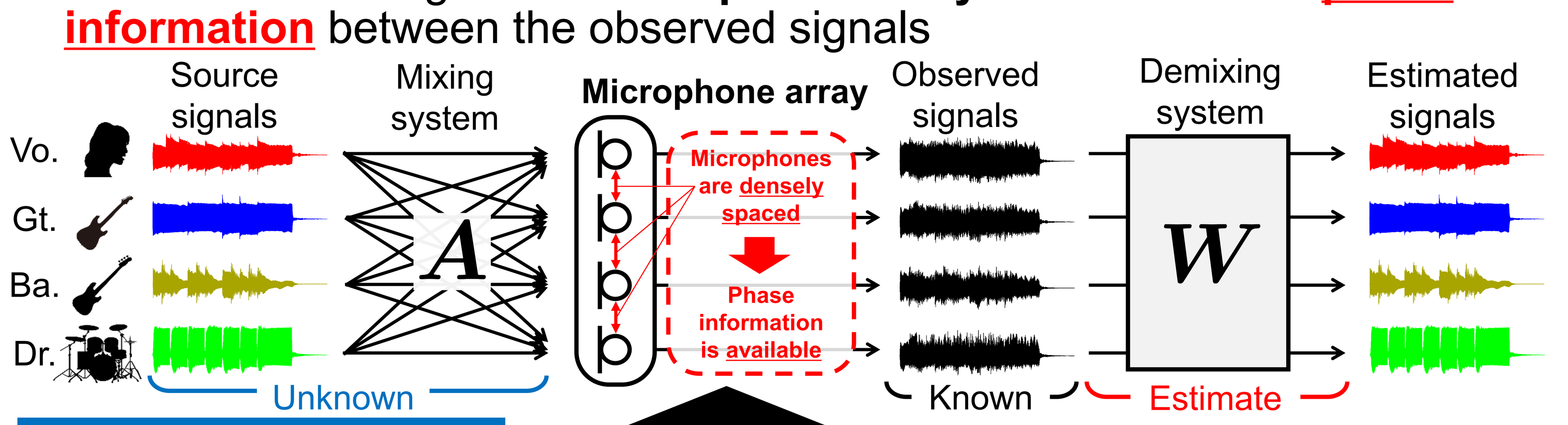


Objective

We address the problem of reducing the bleeding sound in the Vo. microphone

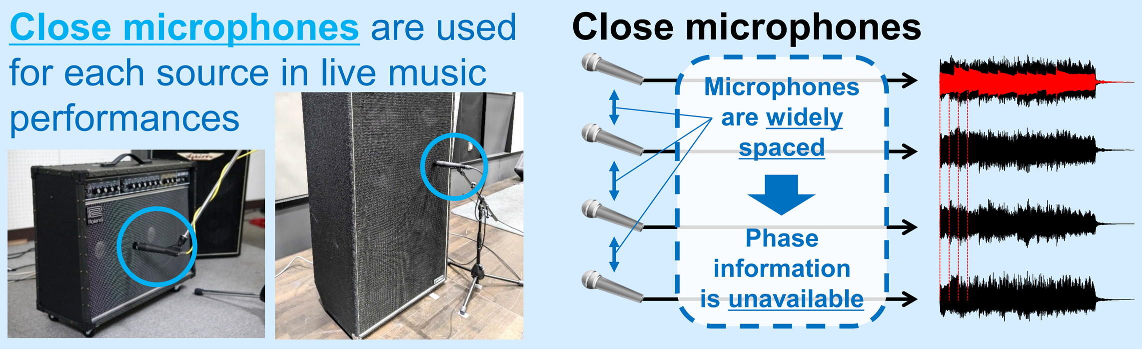
2. Conventional Method

- **Blind source separation (BSS)**
  - estimates **demixing system** W when **mixing system** A is unknown
  - assumes recording with a **microphone array** because it uses **phase information** between the observed signals



Problem of bleeding-sound reduction using BSS

Close microphones are used for each source in live music performances

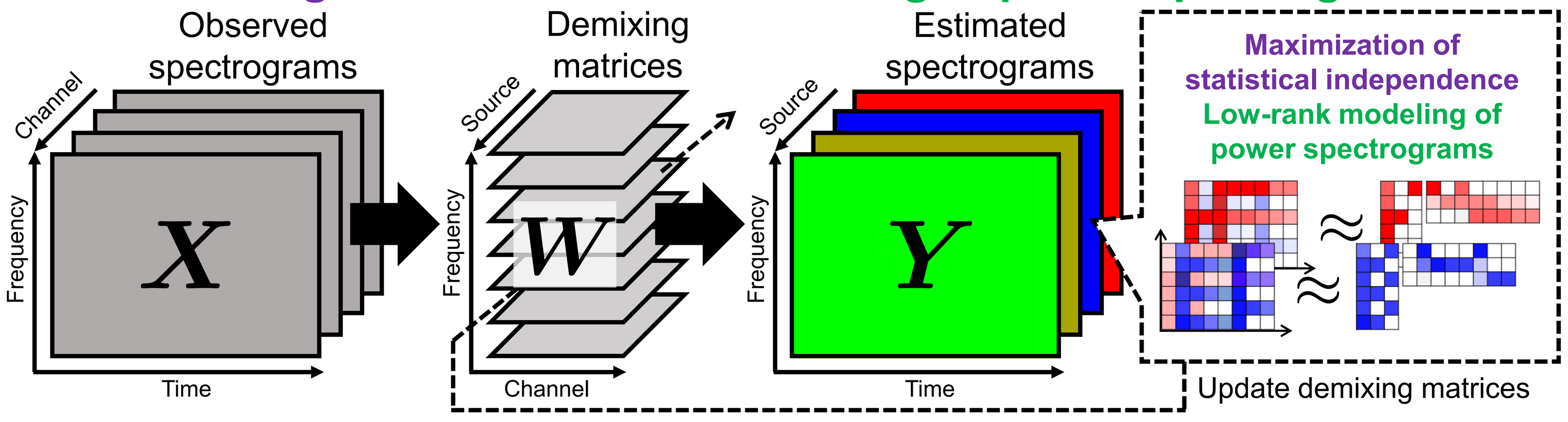


Compared to a microphone array, the spacing between microphones is **wider**

Phase information between the observed signals becomes **unreliable**

Spatial aliasing problem

- **Independent low-rank matrix analysis (ILRMA)** [Kitamura+, 2016]
  - is **BSS** that achieves **maximization of statistical independence between signals** and **low-rank modeling of power spectrograms**



Problem of bleeding-sound reduction using ILRMA

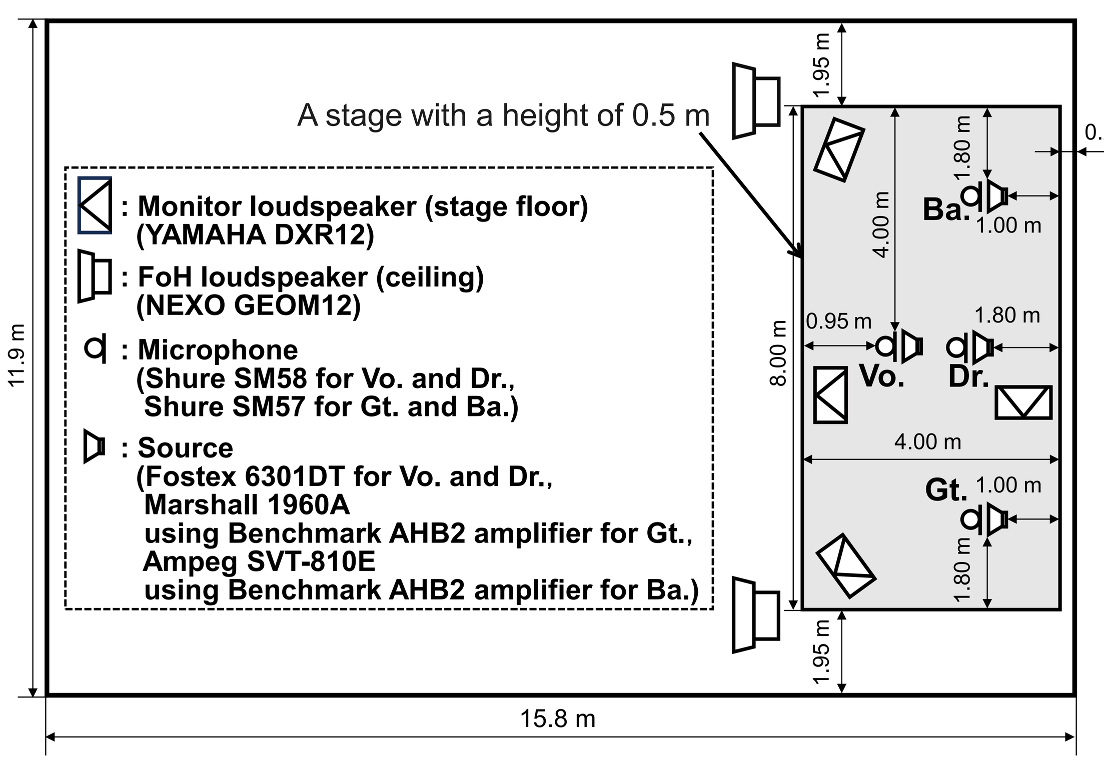
Spatial aliasing degrades the estimation accuracy of demixing matrices

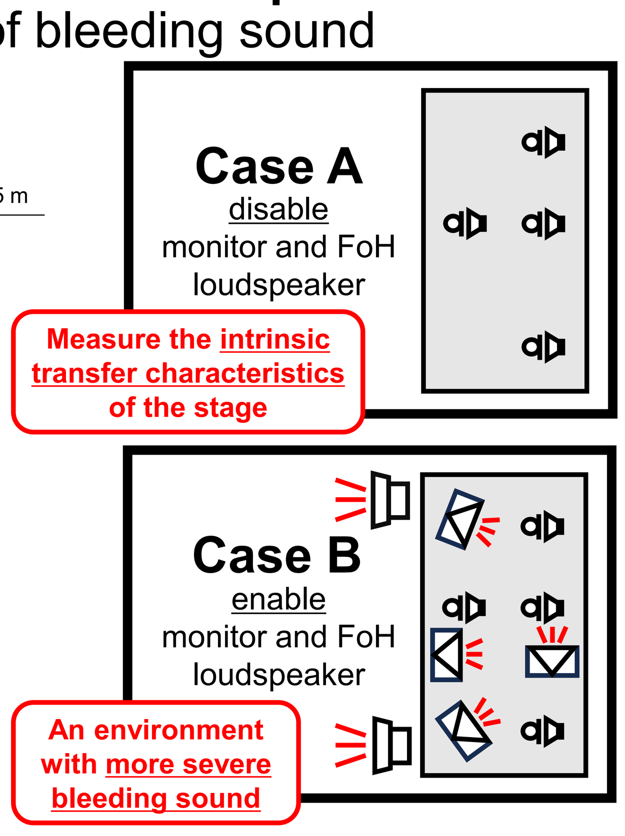
Approach

We propose bleeding-sound reduction for a Vo. microphone that **operates robustly against spatial aliasing** by **modifying ILRMA**

3. Impulse Response Measurement

- **Impulse response measurement in actual live music venue**
  - Set up **two cases of live music venue** and measured **impulse responses** to investigate the characteristics of bleeding sound





- **Measurement result (relative bleeding-sound energy analysis)**
  - Calculated bleeding-sound energy normalized by each target-sound energy

Case A (disable FoH and monitor)

Relative energy [dB]	Vo. source	Gt. source	Ba. source	Dr. source
Vo. microphone	0.0	-17.3	-12.9	1.2
Gt. microphone	-56.3	0.0	-25.3	-29.4
Ba. microphone	-59.7	-34.8	0.0	-31.2
Dr. microphone	-72.7	-44.6	-40.6	0.0

Case B (enable FoH and monitor)

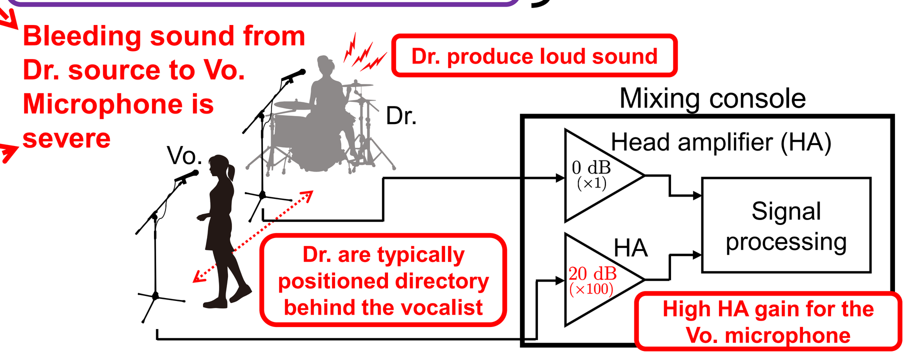
Relative energy [dB]	Vo. source	Gt. source	Ba. source	Dr. source
Vo. microphone	0.0	-9.8	-6.9	1.3
Gt. microphone	-31.5	0.0	-16.9	-19.8
Ba. microphone	-32.5	-21.5	0.0	-23.1
Dr. microphone	-43.9	-34.1	-31.4	0.0

Vo. Microphone contains a large amount of bleeding sound

Other microphones (Gt., Ba., Dr.) contains almost no bleeding sound

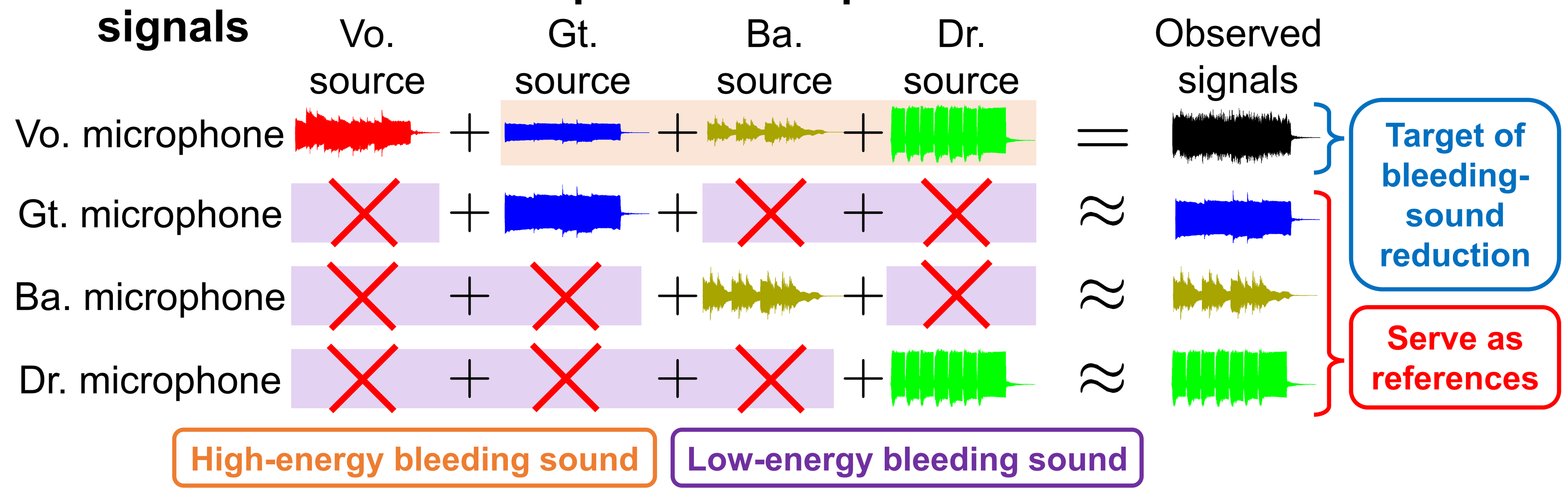
Characteristics of live music venue

Bleeding sound from Dr. source to Vo. Microphone is severe



4. Proposed Method

- **Semi-blind model**
  - treats **low-energy bleeding sound** as **negligible**
  - assumes that **the microphones except for Vo.** can be used as **reference signals**



- **Semi-blind ILRMA (proposed method)**
  - **Introduce the semi-blind model into ILRMA**
  - By introducing reference signals, **the structure of demixing matrices** differs from that of the conventional BSS model

Demixing matrices of original ILRMA

$$W_f = \begin{bmatrix} w_{f11}^* & w_{f12}^* & w_{f13}^* & w_{f14}^* \\ w_{f21}^* & w_{f22}^* & w_{f23}^* & w_{f24}^* \\ w_{f31}^* & w_{f32}^* & w_{f33}^* & w_{f34}^* \\ w_{f41}^* & w_{f42}^* & w_{f43}^* & w_{f44}^* \end{bmatrix} \in \mathbb{C}^{4\times 4}$$

All elements are parameters to be estimated

Demixing matrices of semi-blind ILRMA

$$\tilde{W}_f = \begin{bmatrix} \tilde{w}_{f11}^* & \tilde{w}_{f12}^* & \tilde{w}_{f13}^* & \tilde{w}_{f14}^* \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{C}^{4\times 4}$$

The rows corresponding to the references are fixed to constant value

Advantage

With semi-blind model, the number of parameters are decreased → Robust estimation of $\tilde{W}_f$ / Accurate bleeding-sound reduction

- The model is derived by **substituting the parameter-reduced demixing matrices** into the cost function of ILRMA

Original ILRMA

Cost function

$$\mathcal{J} = -T\sum_f \log|\det W_f|^2 + \sum_{f,t,n}\left[\frac{|w_{fn}^{\mathrm{H}} x_{ft}|^2}{\sum_k b_{fkn} v_{ktn}} + \log\sum_k b_{fkn} v_{ktn}\right]$$

Update rules

$$U_{fn} = \frac{1}{T}\sum_t \frac{1}{\sum_k b_{fkn} v_{ktn}} x_{ft} x_{ft}^{\mathrm{H}}$$

$$w_{fn} \leftarrow (W_f U_{fn})^{-1} e_n$$

$$w_{fn} \leftarrow \frac{w_{fn}}{\sqrt{w_{fn}^{\mathrm{H}} U_{fn} w_{fn}}}$$

$$b_{fkn} \leftarrow b_{fkn}\sqrt{\frac{\sum_t |w_{fn}^{\mathrm{H}} x_{ft}|^2 v_{ktn} \left(\sum_{k'} b_{fk'n} v_{k'tn}\right)^{-2}}{\sum_t v_{ktn}\left(\sum_{k'} b_{fk'n} v_{k'tn}\right)^{-1}}}$$

$$v_{ktn} \leftarrow v_{ktn}\sqrt{\frac{\sum_f |w_{fn}^{\mathrm{H}} x_{ft}|^2 b_{fkn} \left(\sum_{k'} b_{fk'n} v_{k'tn}\right)^{-2}}{\sum_f b_{fkn}\left(\sum_{k'} b_{fk'n} v_{k'tn}\right)^{-1}}}$$

Semi-blind ILRMA

Cost function

$$\mathcal{J} = -T\sum_f \log|\tilde{w}_f^{\mathrm{H}} e_1|^2 + \sum_{f,t}\left[\frac{|\tilde{w}_f^{\mathrm{H}} x_{ft}|^2}{\sum_k \tilde{b}_{fk} \tilde{v}_{kt}} + \log\sum_k \tilde{b}_{fk} \tilde{v}_{kt}\right]$$

Estimation target changes from a matrices to a vectors

Update rules

$$\tilde{U}_f = \frac{1}{T}\sum_t \frac{1}{\sum_k \tilde{b}_{fk} \tilde{v}_{kt}} x_{ft} x_{ft}^{\mathrm{H}}$$

$$\tilde{w}_f \leftarrow \tilde{U}_f^{-1} e_1$$

$$\tilde{w}_f \leftarrow \frac{\tilde{w}_f}{\sqrt{e_1^{\mathrm{H}} \tilde{w}_f}}$$

Using the upper-triangular structure of $\tilde{W}_f$, equations are simplified

$$\tilde{b}_{fk} \leftarrow \tilde{b}_{fk}\sqrt{\frac{\sum_t |\tilde{w}_f^{\mathrm{H}} x_{ft}|^2 \tilde{v}_{kt} \left(\sum_{k'} \tilde{b}_{fk'} \tilde{v}_{k't}\right)^{-2}}{\sum_t \tilde{v}_{kt}\left(\sum_{k'} \tilde{b}_{fk'} \tilde{v}_{k't}\right)^{-1}}}$$

$$\tilde{v}_{kt} \leftarrow \tilde{v}_{kt}\sqrt{\frac{\sum_f |\tilde{w}_f^{\mathrm{H}} x_{ft}|^2 \tilde{b}_{fk} \left(\sum_{k'} \tilde{b}_{fk'} \tilde{v}_{k't}\right)^{-2}}{\sum_f \tilde{b}_{fk}\left(\sum_{k'} \tilde{b}_{fk'} \tilde{v}_{k't}\right)^{-1}}}$$

5. Bleeding-Sound Reduction Experiment

- **Overview**
  - Generated the **observed signals** using **impulse responses measured in actual live music venue**
  - Applied the **following three bleeding-sound reduction methods** and compared the **SDR improvements of the Vo. microphone**

・Gamma-TCNMF [Mizobuchi+, 2025] ・ILRMA ・Semi-blind ILRMA

- **Conditions**
  - **Two types of impulse responses, Case A and Case B,** were used to generate the observed signals
  - **Three different dry sources** were used to generate the observed signals
  - **30 different seeds** were used for the random initialization
  - **Window length** and **three TCNMF hyperparameters** were pre-tuned
  - **The references for the semi-blind ILRMA** were chosen as the **Gt., Ba., and Dr. microphones**

Other parameters
・Window function: Blackman ・Shift size: 1/8 of window length
・Number of iteration: 100 ・Number of basis for ILRMA and Semi-blind ILRMA: 10

- **Result**