# Online Independent Low-Rank Matrix Analysis as a Lightweight and Trainable Model for Real-Time Multichannel Music Source Separation

**Taishi Nakashima** TAIHSI@TMU.AC.JP
*Tokyo Metropolitan University*
*Tokyo, Japan*

**Nobutaka Ono** ONONO@TMU.AC.JP
*Tokyo Metropolitan University*
*Tokyo, Japan*

**Editors:** Tatsuya Komatsu, Keisuke Imoto, Xiaoxue Gao, Nobutaka Ono, Nancy F. Chen

## Abstract

In this paper, we propose an online extension of independent low-rank matrix analysis (ILRMA) for blind music source separation under real-time constraints. Because multitrack stems are rarely released, we target lightweight processing that operates directly on in-the-wild mixtures. The method combines an online Itakura–Saito nonnegative matrix factorization (NMF) update with an online auxiliary-function independent vector analysis (IVA) framework, preserving the low-rank spectral model employed in ILRMA while updating the demixing matrix frame by frame with bounded latency and memory. Simulations on multitrack music mixtures show improved separation accuracy and a real-time factor below one, indicating feasibility for live and interactive scenarios. These results suggest blind separation suitable for low-latency music applications.

**Keywords:** Online blind source separation, independent vector analysis, independent low-rank matrix analysis, nonnegative matrix factorization

## 1. Introduction

Music source separation is increasingly required across a wide range of musical applications Fabbro et al. (2024). These applications include studio postproduction, live sound engineering, automatic transcription, and interactive music systems, all of which depend on accurate separation of multiple simultaneous sources. However, isolated instrument tracks used for training supervised deep neural network (DNN) models are generally inaccessible because multitrack recordings are not released for commercial reasons. As a result, popular systems such as Open-Unmix Stoeter et al. (2019) are typically trained on limited stereo pop datasets and are mostly used for secondary remixing of already-produced material. For genres such as orchestra or chamber ensemble, the scarcity of isolated recordings and the large number of concurrent sources make dataset creation difficult; public datasets remain limited in scale, such as URMP, EnsembleSet, and ChoraleBricks Li et al. (2019); Sarkar et al. (2022); Balke et al. (2025). These constraints motivate the development of lightweight unsupervised and blind algorithms that operate directly on in-the-wild mixtures.

Real-time operation is also essential in live and interactive scenarios. In such settings, long mixtures cannot be buffered and parameter estimation must proceed sequentially with

bounded memory. This requirement naturally leads to online blind source separation, where processing is performed on short audio segments. Independent vector analysis (IVA) Kim et al. (2006); Hiroe (2006); Ono (2011) provides a widely used unsupervised framework for multichannel mixtures, and online auxiliary-function formulations Taniguchi et al. (2014) enable stable real-time updates. However, IVA models only statistical independence and cannot capture the low-rank spectral structure characteristic of musical signals, which limits its performance for music separation.

Unsupervised methods that incorporate explicit spectral modeling are therefore needed. Independent low-rank matrix analysis (ILRMA) Kitamura et al. (2016) addresses this limitation by integrating nonnegative matrix factorization (NMF) into the IVA framework, thereby modeling both independence and low-rank structure. ILRMA is widely regarded as one of the most effective unsupervised and blind approaches for musical mixtures. Its practical use in real-time settings, however, remains limited because ILRMA requires estimating many parameters and its NMF component is sensitive to initialization, sometimes resulting in unstable updates. While ILRMA operates fully unsupervised without pre-mix stems, its NMF bases can optionally be initialized using light genre-specific pretraining, which stabilizes updates while remaining far simpler than DNN training because it estimates only low-dimensional basis matrices rather than paired stems. Despite these advantages, several existing online NMF methods either rely on pre-training or incur computational burdens incompatible with real-time blind processing, as discussed in Section 3.

These observations reveal a clear research gap: ILRMA offers the modeling flexibility required for musical mixtures, yet no existing method achieves ILRMA-level performance in an online, blind, and computationally efficient manner. To address this problem, the present study combines the online NMF update rule based on the Itakura–Saito divergence, which is the source model assumed in ILRMA, with the online auxiliary-function framework of IVA. This approach yields an online extension of ILRMA that preserves low-rank modeling while satisfying the latency and memory constraints required for real-time music source separation. The main contribution of this work is an online ILRMA algorithm that achieves ILRMA-level separation quality under fully blind and low-latency conditions.

## 2. Prior Work and Remaining Gaps

We define online processing as single-pass, frame-by-frame updates with a forgetting factor. Although several methods have been proposed as online NMF, they diverge from this definition: an online natural-gradient (NG) update under the Euclidean distance targets sequential data Zhou et al. (2011); an auxiliary-function approach under the Itakura–Saito (IS) divergence with stabilization procedures has also been studied Lefèvre et al. (2011); multiplicative updates derived via expectation–maximization (EM) for a maximum-likelihood objective have been explored Simon and Vincent (2012); and sequential updates based on the Frobenius norm have been applied to speech enhancement Sack et al. (2022). None of these methods have been evaluated under real-time conditions, and their NMF parameter updates allow multiple passes over the same data; in other words, they revisit past frames to refine factors, which makes strict single-pass real-time operation impractical.

Approaches that combine online auxiliary-function-based IVA (AuxIVA) with pretrained NMF bases have also been explored Wang et al. (2021), but they assume a spherically

symmetric Laplace source model while deriving IS-based updates and still require speech-domain pretraining. Consequently, no prior work simultaneously meets all requirements for in-the-wild mixtures: multichannel processing, the IS divergence objective, auxiliary-function updates, no pretraining, and single-pass operation. To date, no online BSS method has been proposed that satisfies these conditions.

## 3. Conventional Methods

### 3.1. Signal Model

First, we describe the observation model used in this study. Let the number of sources and microphones be equal to $K$. In the short-time Fourier transform domain, we assume that the observed signal $\boldsymbol{x}_{f,t}$ is a linear mixture of $K$ sources:

$$\boldsymbol{x}_{f,t} = \sum_{k=1}^{K} \boldsymbol{a}_{k,f,t} s_{k,f,t} \in \mathbb{C}^K. \tag{1}$$

Here, $\boldsymbol{a}_{k,f,t}$ is the steering vector of source $k$, $s_{k,f,t}$ is the source signal, $f = 1, \ldots, F$ is the frequency-bin index, and $t = 1, \ldots, T$ is the time-frame index. Because we address online BSS in this work, note that the steering vectors are time-varying. The goal of online BSS is to estimate parameters called *demixing matrix* $W_{f,t} \in \mathbb{C}^{K \times K}$ that recover the source signals when only the current or past observed signals are available. The estimated signals are computed as $\boldsymbol{y}_{f,t} = W_{f,t}\boldsymbol{x}_{f,t}$, where $\boldsymbol{w}_{k,f,t}^{\mathsf{H}} \in \mathbb{C}^K$ is the $k$-th row of $W_{f,t}$ and called demixing vector.

### 3.2. Auxiliary-function-based Independent Vector Analysis

Before explaining online IVA, we briefly outline batch AuxIVA Ono (2011). In this subsection, the demixing matrix has no time-frame index $t$ since batch processing assumes a time-invariant system. AuxIVA minimizes the following objective function Ono (2011):

$$r_{k,t} = \sqrt{\sum_{f=1}^{F} |\boldsymbol{w}_{k,f,t}^{\mathsf{H}} \boldsymbol{x}_{f,t}|^2}, \; V_{k,f} = \frac{1}{T} \sum_{t=1}^{T} \varphi(r_{k,t}) \boldsymbol{x}_{f,t} \boldsymbol{x}_{f,t}^{\mathsf{H}}, \tag{2}$$

$$J^{+} = \sum_{f=1}^{F} \left( \sum_{k=1}^{K} \boldsymbol{w}_{k,f}^{\mathsf{H}} V_{k,f} \boldsymbol{w}_{k,f} - \log |\det W_f|^2 \right). \tag{3}$$

Here, $\varphi$ is defined by $\varphi(r) = \frac{\psi'(r)}{2r}$ using the first derivative $\psi'$ of the contrast $\psi$. The contrast function $\psi$ is determined by the prior distribution assumed for the source signals. Unless otherwise stated, we set $\varphi(r) = \frac{F}{r^2}$, which corresponds to assuming a time-varying variance complex Gaussian distribution as the source model Ono (2012). By minimizing the above objective, update rules such as iterative projection Ono (2011) and iterative source steering Scheibler and Ono (2020) are obtained. $V_k$ is an intermediate variable called the *weighted covariance matrix*, and it requires the observed signals for all time frames $t = 1, \ldots, T$.

### 3.3. Online Auxiliary-function-based Independent Vector Analysis

Online AuxIVA estimates the demixing matrix $W_{f,t}$ for each frame $t$ by minimizing the following function Taniguchi et al. (2014); Nakashima et al. (2023):

$$\hat{y}_{k,f,t} = \boldsymbol{w}_{k,f,t-1}^{\mathsf{H}} \boldsymbol{x}_{f,t},\ r_{k,t} = \sqrt{\sum_{f=1}^{F} |\hat{y}_{k,f,t}|}, \tag{4}$$

$$V_{k,f,t} = \alpha V_{k,f,t-1} + (1-\alpha)\varphi(r_{k,t})\boldsymbol{x}_{f,t}\boldsymbol{x}_{f,t}^{\mathsf{H}}, \tag{5}$$

$$J_t^{+} = \sum_{f=1}^{F} \left( \sum_{k=1}^{K} \boldsymbol{w}_{k,f,t}^{\mathsf{H}} V_{k,f,t} \boldsymbol{w}_{k,f,t} - \log|\det W_{f,t}|^2 \right), \tag{6}$$

where, $0 < \alpha \leq 1$ is a *forgetting factor*. This can be interpreted as an approximation of the objective using the instantaneous observed signals, allowing the batch AuxIVA update rules to be used with the same formulas in the online case.

### 3.4. Independent Low-Rank Matrix Analysis

ILRMA can be regarded as an extension of AuxIVA that assumes the complex spectrograms of source signals are generated from zero-mean circular complex Gaussian distributions. ILRMA minimizes the following objective function Kitamura et al. (2016):

$$r_{k,f,t} = \sum_{\ell=1}^{L} b_{k,f,\ell} c_{k,\ell,t},\quad V_{k,f} = \frac{1}{T} \sum_{t=1}^{T} \varphi(r_{k,f,t}) \boldsymbol{x}_{f,t} \boldsymbol{x}_{f,t}^{\mathsf{H}}, \tag{7}$$

$$J^{+} = \sum_{f=1}^{F} \Bigg( \sum_{k=1}^{K} \left( \boldsymbol{w}_{k,f}^{\mathsf{H}} V_{k,f} \boldsymbol{w}_{k,f} - \frac{1}{T} \log r_{k,f,t} \right) - \log|\det W_f|^2 \Bigg). \tag{8}$$

Here, $\ell = 1, \ldots, L$ is the basis index, and $L$ is the number of bases, which is a user-specified parameter. $b_{k,f,\ell}$ and $c_{k,\ell,t}$ are nonnegative real numbers used to obtain a low-rank approximation of the amplitude of the $k$-th separated signal $y_{k,f,t}$, and they are called the *basis* and *activation*, respectively. $b_{k,f,\ell}$ and $c_{k,\ell,t}$ are updated by the following multiplicative updates Kitamura et al. (2016):

$$b_{k,f,\ell} \leftarrow b_{k,f,\ell} \left( \frac{\sum_t |y_{k,f,t}|^2 c_{k,\ell,t} (\sum_i b_{k,f,i} c_{k,i,t})^{-2}}{\sum_t c_{k,\ell,t} (\sum_i b_{k,f,i} c_{k,i,t})^{-1}} \right), \tag{9}$$

$$c_{k,\ell,t} \leftarrow c_{k,\ell,t} \left( \frac{\sum_f |y_{k,f,t}|^2 b_{k,f,\ell} (\sum_i b_{k,f,i} c_{k,i,t})^{-2}}{\sum_f b_{k,f,\ell} (\sum_i b_{k,f,i} c_{k,i,t})^{-1}} \right). \tag{10}$$

The update rules (9), (10) for the ILRMA source model are equivalent to those of NMF based on the Itakura–Saito divergence (Itakura–Saito NMF; ISNMF) Févotte et al. (2009). In ILRMA, this is equivalent to estimating two kinds of nonnegative real parameters that approximate the power $|y_{k,f,t}|^2$ of each separated signal $y_{k,f,t}$. Because ISNMF is a batch method, it requires parameters for all time frames $t = 1, \ldots, T$. In particular, as shown in (9), not only the weighted covariance matrices but also the bases depend on all time frames.

### 3.5. Online Nonnegative Matrix Factorization based on Itakura–Saito Divergence

Online ISNMF sequentially estimates the NMF parameters at each time frame $t$ Lefèvre et al. (2011). Assume $K = 1$ and define the spectrogram matrix $X \in \mathbb{C}^{F\times T}$ with $(f,t)$-th element $x_{f,t}$; each column $\boldsymbol{x}_t$ stacks all frequencies at frame $t$. Online ISNMF estimates $B \in \mathbb{R}_+^{F\times L}$ and $C \coloneqq \begin{bmatrix}\boldsymbol{c}_1 & \dots & \boldsymbol{c}_T\end{bmatrix} \in \mathbb{R}_+^{L\times T}$ that approximate the power $|X|^2$ of the observed spectrogram. Absolute value, powers, divisions, and $\odot$ denote elementwise operations, and $\mathbb{R}_+$ denotes the nonnegative reals. The forgetting factor for the source-model parameters is denoted by $\beta$, and the mini-batch length is denoted by $N$. We define $\rho \coloneqq \beta^{N/T}$ following the same formulation as Lefèvre et al. (2011). The update rule can be written as the following single-pass updates for $t = 1, \dots, T$:

$$\boldsymbol{c}_t \leftarrow \boldsymbol{c}_{t-1} \odot \frac{B^{\mathsf{T}}\left(|\boldsymbol{x}_t|^2 \odot (B\boldsymbol{c}_{t-1})^{-2}\right)}{B^{\mathsf{T}}(B\boldsymbol{c}_{t-1})^{-1}}, \tag{11}$$

$$\tilde{P} \leftarrow \left(\frac{|\boldsymbol{x}_t|^2}{(B\boldsymbol{c}_t)^2}\boldsymbol{c}_t^{\mathsf{T}}\right) \odot B^2, \quad P_t \leftarrow \rho\, P_{t-1} + \tilde{P}, \tag{12}$$

$$\tilde{Q} \leftarrow \frac{1}{B\boldsymbol{c}_t}\boldsymbol{c}_t^{\mathsf{T}}, \quad Q_t \leftarrow \rho\, Q_{t-1} + \tilde{Q}, \tag{13}$$

$$B \leftarrow \left(\frac{P_t}{Q_t}\right)^{\frac{1}{2}}. \tag{14}$$

Here, $N$ controls how often the basis matrix $B$ (called the "dictionary" in the original paper) is updated (mini-batch length). Revisiting the same mixture requires explicitly cycling through the data, because $N > 1$ only delays updates and does not create additional passes.

## 4. Proposed Method: Online Independent Low-Rank Matrix Analysis

As discussed in Section 3, online BSS algorithms have been realized by approximating parameters that require all time frames $t = 1, \dots, T$ with framewise computations. Since batch ILRMA can be interpreted as combining AuxIVA and ISNMF, the source-model parameter updates in the online setting should also be based on the IS divergence. We therefore attempt to realize online ILRMA by combining online AuxIVA and online ISNMF. Because the updates in (11)–(14) provide a low-rank approximation for a *single* source, we replace the observed power $|\boldsymbol{x}_t|^2$ with each separated power $|\hat{\boldsymbol{y}}_{k,t}|^2$ and apply the NMF updates to every source in parallel. This follows the same manner as the batch extension from AuxIVA to ILRMA, where source-wise variance models are injected while keeping the demixing updates unchanged. To satisfy real-time constraints, we couple the forgetting factors of online AuxIVA and online ISNMF so that both spatial and spectral parameters are updated in a single pass without revisiting past frames. In the online ISNMF part, the intermediate accumulators $P$ and $Q$ for updating the basis matrices $B$ are accumulated with coefficient 1 at each frame and decayed only when the mini-batch length $N$ is reached via $\rho = \alpha^{N/\max(u,1)}$, where $u$ counts how many times $P, Q$ have been updated. This matches the decay scheduling in our implementation and approximates per-frame forgetting with

**Algorithm 1** Online ILRMA

Initialize $W_f$, $V_{k,f,0}$, $B_k$, $\boldsymbol{c}_{k,0}$, $P_{k,0}$, $Q_{k,0}$; set $u \leftarrow 0$, $j \leftarrow 0$
**for** $t = 1, \ldots, T$ **do**
  $\hat{\boldsymbol{y}}_{f,t} \leftarrow W_{f,t-1}\boldsymbol{x}_{f,t}$
  **for** $i = 1, \ldots, N_{\mathrm{i}}$ **do**
    **for** $k = 1, \ldots, K$ **do**
      $\boldsymbol{c}_{k,t} \leftarrow \boldsymbol{c}_{k,t-1} \odot \dfrac{B_k^{\mathsf{T}}\left(|\hat{\boldsymbol{y}}_{k,t}|^2 \oslash (B_k\boldsymbol{c}_{k,t-1})^2\right)}{B_k^{\mathsf{T}}\left(\mathbf{1} \oslash (B_k\boldsymbol{c}_{k,t-1})\right)}$
      $\boldsymbol{r}_{k,t} \leftarrow B_k\boldsymbol{c}_{k,t} + \epsilon$
      $\tilde{P}_k \leftarrow \left(|\hat{\boldsymbol{y}}_{k,t}|^2 \oslash \boldsymbol{r}_{k,t}^2\right)\boldsymbol{c}_{k,t}^{\mathsf{T}} \odot B_k^2$
      $\tilde{Q}_k \leftarrow (\mathbf{1} \oslash \boldsymbol{r}_{k,t})\,\boldsymbol{c}_{k,t}^{\mathsf{T}}$
      $P_k \leftarrow P_k + \tilde{P}_k, \quad Q_k \leftarrow Q_k + \tilde{Q}_k$
      $u \leftarrow u + 1, \quad j \leftarrow j + 1$
      $V_{k,f,t} \leftarrow \alpha V_{k,f,t} + (1-\alpha)\,\varphi(\boldsymbol{r}_{k,t})\,\boldsymbol{x}_{f,t}\boldsymbol{x}_{f,t}^{\mathsf{H}}$
    **end for**
    **if** $j \geq N$ **then**
      $\rho \leftarrow \alpha^{N/\max(u,1)}$
      **for** $k = 1, \ldots, K$ **do**
        $P_k \leftarrow \rho\,P_k, \quad Q_k \leftarrow \rho\,Q_k$
        $B_k \leftarrow \left(\dfrac{P_k}{Q_k}\right)^{\frac{1}{2}}$
        $\bar{\boldsymbol{b}} \leftarrow \mathbf{1}^{\mathsf{T}}B_k$
        $B_k \leftarrow B_k \oslash \left(\mathbf{1}\,\bar{\boldsymbol{b}}\right)$
        $P_k \leftarrow P_k \oslash \left(\mathbf{1}\,\bar{\boldsymbol{b}}\right), \quad Q_k \leftarrow Q_k \odot \left(\mathbf{1}\,\bar{\boldsymbol{b}}\right)$
      **end for**
      $j \leftarrow 0$
    **end if**
    **for** $k = 1, \ldots, K$ **do**
      $\boldsymbol{w}_{k,f,t} \leftarrow (W_{f,t}V_{k,f,t})^{-1}\,\boldsymbol{e}_k$
      $\boldsymbol{w}_{k,f,t} \leftarrow \boldsymbol{w}_{k,f,t}\,\left(\boldsymbol{w}_{k,f,t}^{\mathsf{H}}V_{k,f,t}\boldsymbol{w}_{k,f,t}\right)^{-1/2}$
    **end for**
  **end for**
  $\boldsymbol{y}_{f,t} \leftarrow W_{f,t}\boldsymbol{x}_{f,t}$
**end for**

fewer exponentiations. We also adopt the column-sum normalization ("rescaling" as in the original paper) for $B$ after each mini-batch update to improve the numerical stability of the online basis updates. The per-frame computational complexity of the proposed online ILRMA is $O(FK^4 + FKL)$: $O(FK^4)$ from the IP-based demixing updates and $O(FKL)$ from the IS-NMF updates. Because prior work (e.g., Kitamura et al. (2016)) reports limited gains beyond about ten bases, the increase in computation is typically dominated by the number of microphones/sources rather than by the mini-batch length $N$. We use $\odot$ and $\oslash$ for element-wise product and division, respectively, and $\mathbf{1}$ for an all-ones matrix of appropriate shape. The resulting algorithm is summarized in Algorithm 1.

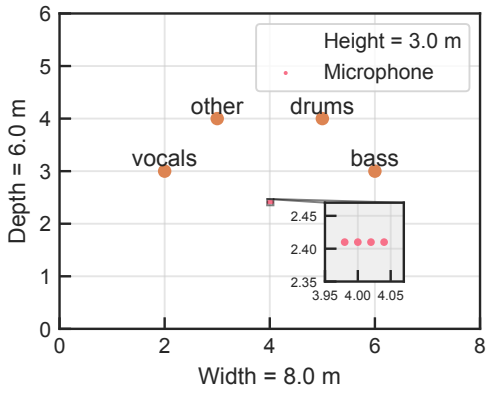


Figure 1: Room geometry and microphone/source positions.

## 5. Experiments

### 5.1. Setup

We conducted simulation experiments to verify the efficacy of the proposed method. Mixtures were generated by convolving source signals with room impulse responses synthesized by the image source method implemented in pyroomacoustics Scheibler et al. (2018). As source signals, we used the four stems, `bass`, `drums`, `others`, and `vocals` from MUSDB18HQ dataset Rafii et al. (2019), selecting 100 tracks from the training set. The original sampling rate is 44.1 kHz; we downsampled to 16 kHz to reduce computational load and used 30-second segments where the sources are active. A 4-channel linear microphone array with 2 cm spacing was placed at the center of the room, and the reverberation time was set to about 200 ms. Room dimensions were 8.0 m × 6.0 m × 3.0 m. A linear microphone array with four microphones and spacing 0.02 m was centered at (4.01 m, 2.41 m, 1.7 m). Source positions were bass (6.0 m, 3.0 m, 1.7 m), drums (5.0 m, 4.0 m, 1.7 m), other (3.0 m, 4.0 m, 1.7 m), and vocals (2.0 m, 3.0 m, 1.7 m) as shown in 1. The same spatial configuration was used for all tracks. For the short-time Fourier transform, we used a window length of 2048, a hop size of 512, and a Hann analysis window.

As baselines, we used Batch AuxIVA, Batch ILRMA, Blockbatch AuxIVA, Blockbatch ILRMA, and Online AuxIVA. And we compared with our proposed Online ILRMA. Online AuxIVA used a forgetting factor $\alpha$ of 0.99, and online ISNMF used a mini-batch length $N$ of 2. For ILRMA in each method above, the number of bases $L$ was set to 10. In Blockbatch AuxIVA and Blockbatch ILRMA, each method was applied in batch to every 200 frames of the observed signals after the short-time Fourier transform. Separation performance was evaluated by the improvement in scale-invariant signal-to-distortion ratio (SI-SDRi) Le Roux et al. (2019).

### 5.2. Results and discussions

Table 1 shows the SI-SDRi for each method and stem. Online ILRMA outperformed Online AuxIVA for all stems; the gain was largest for drums (+3.2 dB) and notable for bass (+2.0 dB) and other (+1.5 dB). Both online methods remained below the batch counterparts because the single-pass constraint and forgetting ($\alpha = 0.99$) limit late-frame refinement,

| Method | bass | drums | other | vocals |
|---|---|---|---|---|
| Batch AuxIVA | −1.22 | 7.72 | 3.19 | 6.72 |
| Batch ILRMA | −0.838 | 8.45 | 4.40 | 6.65 |
| Blockbatch AuxIVA | −3.69 | 4.24 | 0.365 | 2.21 |
| Blockbatch ILRMA | −3.37 | 3.28 | 0.311 | 1.78 |
| Online AuxIVA | −5.08 | 1.20 | −1.08 | 1.25 |
| Online ILRMA | −3.09 | 4.40 | 0.370 | 1.51 |

Table 1: SI-SDRi by stems and methods.

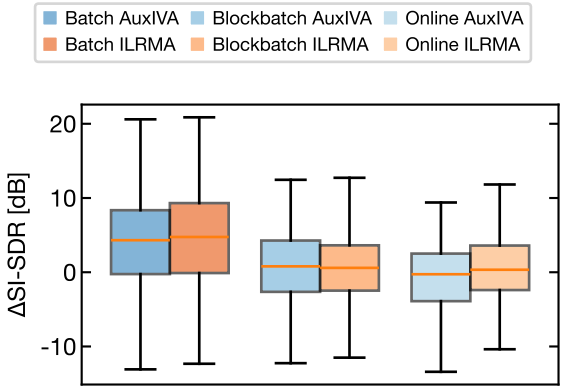


Figure 2: Box plots of SI-SDR improvements across stems.

whereas batch ILRMA can re-estimate bases with all frames. Block-batch results lay between batch and online, reflecting partial access to future frames but still lacking full passes. The performance gap between online and batch methods is attributed to the limited temporal context of online updates: statistics such as the weighted covariance matrices cannot be refined by revisiting past frames, making them less accurate than those of batch and block-batch processing. Even with this limitation, Online ILRMA attains SI-SDRi close to block-batch results without any backward passes, highlighting its efficiency under streaming constraints.

Figure 2 shows box plots of SI-SDR improvements across all stems and tracks, and Figure 3 illustrates their temporal variations. Online ILRMA achieved a slightly higher median than Online AuxIVA and converged faster; both methods reached comparable steady-state SI-SDRi, but Online ILRMA arrives there more quickly. Occasional numerical instabilities produced outliers for Online ILRMA, and vocals and drums showed relatively higher performance, whereas bass was lowest for all methods, likely because bass energy is narrowband and correlated with drums.

Figure 4 shows the real-time factor (RTF) of the block-batch and online methods. All methods achieved RTFs below 1, indicating real-time capability. Although the proposed Online ILRMA requires slightly more computation than Online AuxIVA due to the NMF updates, the increase is marginal. Under our setting ($K = 4$, $F = 1025$, $L = 10$, $N = 2$, $N_{\mathrm{i}} = 1$), the per-frame cost is on the order of $3 \times 10^5$ FLOP, and the trainable parameters

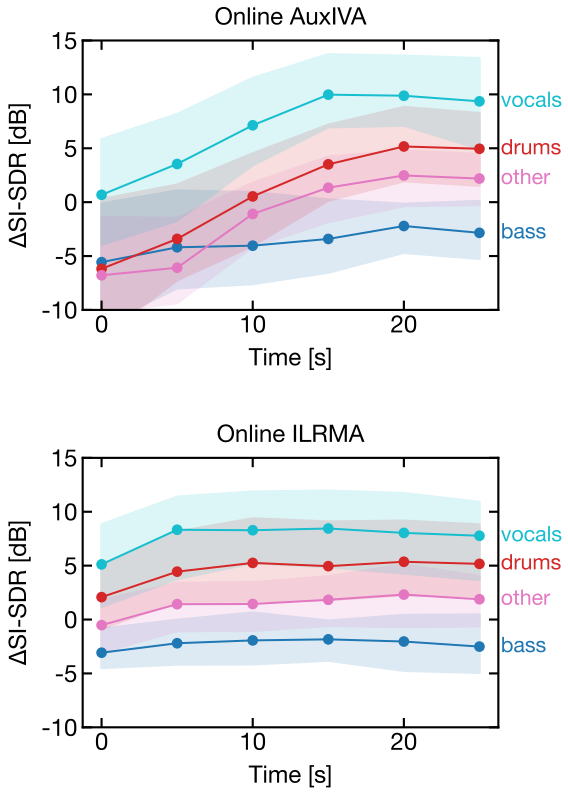


Figure 3: Temporal variations of SI-SDR improvements of conventional Online AuxIVA and proposed Online ILRMA.

are about $5.7 \times 10^4$ (demixing + bases). In contrast, the 4-stem Spleeter U-Net uses 6 down/6 up convolutions with tens of millions of parameters (about 35M parameters with a model size of roughly 140 MB Hennequin et al. (2020)), yet our Online ILRMA keeps RTF $<1$ on CPU without such large models. For reference, Spleeter can run about 100× faster than real time on a GTX1080-class GPU Hennequin et al. (2020).

## 6. Conclusion

This paper realized an online version of independent low-rank matrix analysis by combining online auxiliary-function-based independent vector analysis with online nonnegative matrix factorization based on the Itakura–Saito divergence. This enables sequential estimation of the demixing matrix without storing all parameters, even when batch processing is infeasible due to a large number of microphones or long signals. Simulation results showed that the proposed method improves separation performance even for musical signals that have been difficult to separate with conventional methods. Future work includes simulations on larger datasets, comprehensive ablations over forgetting factors, block sizes, and number of bases, and implementation of real-time processing. The proposed method keeps the demixing matrix and NMF bases trainable and updates them on the fly as new audio arrives, without

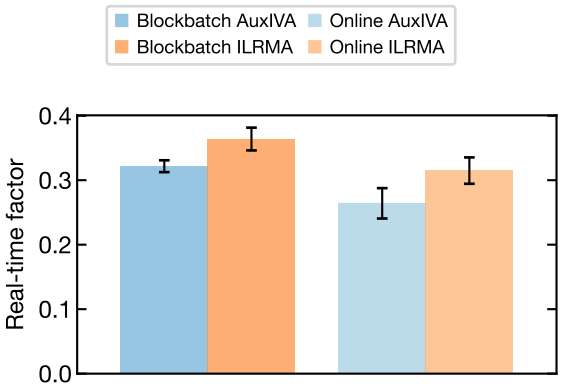


Figure 4: Real-time factor of each method.

relying on pre-trained models or large buffers, thereby matching the "lightweight and trainable" objective stated in the title. The proposed online ILRMA also remains far lighter than DNN-based music source separation: its trainable parameters (demixing matrices and NMF bases) are far fewer than typical U-Net-based separators, and the experimental results indicate promising feasibility for real-time processing without large models.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP24K23896 and JP25H01150, Japan.

## References

Stefan Balke, Axel Berndt, and Meinard Müller. Choralebricks: A modular multitrack dataset for wind music research. *Transactions of the International Society for Music Information Retrieval*, 8(1):39–54, 2025. doi: 10.5334/tismir.252.

Giorgio Fabbro, Stefan Uhlich, Chieh-Hsin Lai, Woosung Choi, Marco Martínez-Ramírez, Weihsiang Liao, Igor Gadelha, Geraldo Ramos, Eddie Hsu, Hugo Rodrigues, Fabian-Robert Stöter, Alexandre Défossez, Yi Luo, Jianwei Yu, Dipam Chakraborty, Sharada Mohanty, Roman Solovyev, Alexander Stempkovskiy, Tatiana Habruseva, Nabarun Goswami, Tatsuya Harada, Minseok Kim, Jun Hyung Lee, Yuanliang Dong, Xinran Zhang, Jiafeng Liu, and Yuki Mitsufuji. The sound demixing challenge 2023 – music demixing track. *Transactions of the International Society for Music Information Retrieval*, 7(1):63–84, 2024. doi: 10.5334/tismir.171.

Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu. Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830, 2009. doi: 10.1162/neco.2008.04-08-771.

Romain Hennequin, Anis Khlif, Felix Voituret, and Manuel Moussallam. Spleeter: a fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, 5(50):2154, 2020. doi: 10.21105/joss.02154.

Atsuo Hiroe. Solution of permutation problem in frequency domain ICA, using multivariate probability density functions. In *Proc. ICA*, pages 601–608, March 2006.

Taesu Kim, Hagai T Attias, Soo-Young Lee, and Te-Won Lee. Blind source separation exploiting higher-order frequency dependencies. *IEEE Audio, Speech, Language Process.*, 15(1):70–79, January 2006. doi: 10.1109/TASL.2006.872618.

Daichi Kitamura, Nobutaka Ono, Hiroshi Sawada, Hirokazu Kameoka, and Hiroshi Saruwatari. Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(9):1622–1637, September 2016. doi: 10.1109/TASLP.2016.2577880.

Jonathan Le Roux, Scott Wisdom, Hakan Erdogan, and John R. Hershey. SDR — half-baked or well done? In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 626–630, May 2019. doi: 10.1109/ICASSP.2019.8683855.

Augustin Lefèvre, Francis Bach, and Cedric Fevotte. Online algorithms for nonnegative matrix factorization with the itakura-saito divergence. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 313–316, 2011. doi: 10.1109/ASPAA.2011.6082314.

Bochen Li, Xinzhao Liu, Karthik Dinesh, Zhiyao Duan, and Gaurav Sharma. Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. *IEEE Transactions on Multimedia*, 21(2):522–535, 2019. doi: 10.1109/TMM.2018.2856090.

Taishi Nakashima, Rintaro Ikeshita, Nobutaka Ono, Shoko Araki, and Tomohiro Nakatani. Fast online source steering algorithm for tracking single moving source using online independent vector analysis. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, June 2023. doi: 10.1109/ICASSP49357.2023.10094962.

Nobutaka Ono. Stable and fast update rules for independent vector analysis based on auxiliary function technique. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 189–192, October 2011. doi: 10.1109/ASPAA.2011.6082320.

Nobutaka Ono. Auxiliary-function based independent vector analysis with power of vector-norm type weighting functions. In *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–4, December 2012.

Zafar Rafii, Antoine Liutkus, Fabian-Robert Stöter, Stylianos Ioannis Mimilakis, and Rachel Bittner. MUSDB18-HQ - an uncompressed version of MUSDB18, August 2019. URL https://doi.org/10.5281/zenodo.3338373.

Andrew Sack, Wenzhao Jiang, Michael Perlmutter, Palina Salanevich, and Deanna Needell. On audio enhancement via online non-negative matrix factorization. In *2022 56th Annual Conference on Information Sciences and Systems (CISS)*, pages 287–291, 2022. doi: 10.1109/CISS53076.2022.9751157.

Saurjya Sarkar, Emmanouil Benetos, and Mark Sandler. EnsembleSet: a new high quality synthesised dataset for chamber ensemble separation. In *Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR 2022)*, pages 625–632, 2022. doi: 10.5281/zenodo.7316740.

Robin Scheibler and Nobutaka Ono. Fast and stable blind source separation with rank-1 updates. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 236–240, 2020.

Robin Scheibler, Eric Bezzam, and Ivan Dokmanić. Pyroomacoustics: A Python package for audio room simulation and array processing algorithms. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 351–355, April 2018. doi: 10.1109/ICASSP.2018.8461310.

Laurent S. R. Simon and Emmanuel Vincent. A general framework for online audio source separation. In *Latent Variable Analysis and Signal Separation*, volume 7191, pages 397–404, 2012. doi: 10.1007/978-3-642-28551-6_49.

Fabian-Robert Stoeter, Stefan Uhlich, Antoine Liutkus, and Yuki Mitsufuji. Open-unmix – a reference implementation for music source separation. *Journal of Open Source Software*, 4(41):1667, 2019. doi: 10.21105/joss.01667.

Toru Taniguchi, Nobutaka Ono, Akinori Kawamura, and Shigeki Sagayama. An auxiliary-function approach to online independent vector analysis for real-time blind source separation. In *Proceedings of Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, pages 107–111, May 2014.

Taihui Wang, Feiran Yang, Rui Zhu, and Jun Yang. Real-time independent vector analysis using semi-supervised nonnegative matrix factorization as a source model. In *Proc. Interspeech*, pages 1842–1846, August 2021. doi: 10.21437/Interspeech.2021-146.

Guoxu Zhou, Zuyuan Yang, Shengli Xie, and Jun-Mei Yang. Online blind source separation using incremental nonnegative matrix factorization with volume constraint. 22(4):550–560, 2011. doi: 10.1109/TNN.2011.2109396.