Optimal divergence diversity for superresolution-based nonnegative matrix factorization *

◎ Daichi Kitamura, Hiroshi Saruwatari, Satoshi Nakamura (Nara Institute of Science and Technology)
Kazunobu Kondo, Yu Takahashi (Yamaha Corporation), Hirokazu Kameoka (The University of Tokyo)

1 Introduction

In recent years, music and acoustic signal separation based on nonnegative matrix factorization (NMF) [1] has been a very active area of signal-processing research. NMF of acoustic signals decomposes an input spectrogram into the product of a spectral basis matrix and its activation matrix. In particular, supervised NMF (SNMF) [2], which includes a priori training with some sample sounds of a target instrument, can extract the target signal to some extent. However, in the case of a mixture consisting of many sources, the source extraction performance is markedly degraded when only single-channel observation is available.

Multichannel NMF, which is a natural extension of NMF to a multichannel signal, has been proposed as an unsupervised method [3]. However, such an unsupervised separation is a difficult problem because the decomposition is underspecified. Hence, algorithms used for multichannel NMF have strong dependence on initial values and lack robustness.

As another means of addressing multichannel signal separation, a hybrid method, which concatenates superresolution-based SNMF after directional clustering, has been proposed by the authors [4, 5]. This method uses index information generated by binary masking of directional clustering so that spectral chasms can be regarded as *unseen* observations, and finally reconstructs the target source components via spectrogram extrapolation using the supervised bases as a *dictionary*. Also, we have proposed some update algorithms for the superresolution-based SNMF based on β-divergence, which includes Itakura-Saito divergence (*IS-divergence*), generalized Kullback-Leibler divergence, (*KL-divergence*) and Euclidean distance (*EUC-distance*) [5]. In general SNMF-based music signal separation, KL-divergence is often used as a cost function because the spectrogram of the music signals tends to become sparse, and KL-divergence-based SNMF fits to represent such sparse signals [6]. However, it has been experimentally confirmed that the optimal divergence for superresolution-based SNMF is EUC-distance [5]. This difference of performance between EUC-distance and KL-divergence is due to the difference of a spatial condition in each source, and the optimal divergence temporally fluctuates because the spatial condition is not consistent in the general music signal. Therefore, these divergence should be changed in each time frame automatically. To solve this problem, in this paper, we propose a new scheme for frame-wise divergence diversity to separate the target signal using optimal divergence.

2 Conventional Method

2.1 SNMF

In SNMF, a priori spectral patterns (bases) should be trained in advance as a basis dictionary. The following equation represents the decomposition in SNMF;

$$\boldsymbol{Y} \simeq \boldsymbol{F}\boldsymbol{G} + \boldsymbol{H}\boldsymbol{U}, \tag{1}$$

where $\boldsymbol{Y}(\in \mathbb{R}_{\geq 0}^{\Omega \times T})$ is an observed spectrogram, $\boldsymbol{F}(\in \mathbb{R}_{\geq 0}^{\Omega \times K})$ is a supervised basis matrix trained in advance, which includes spectral patterns of the target signal as column vectors, $\boldsymbol{G}(\in \mathbb{R}_{\geq 0}^{K \times T})$ is the activation matrix that corresponds to $\boldsymbol{F}$, $\boldsymbol{H}(\in \mathbb{R}_{\geq 0}^{\Omega \times L})$ represents the residual spectral patterns that cannot be expressed by $\boldsymbol{F}\boldsymbol{G}$, and $\boldsymbol{U}(\in \mathbb{R}_{\geq 0}^{L \times T})$ is the activation matrix that corresponds to $\boldsymbol{H}$. Moreover, Ω is the number of frequency bins, T is the number of frames of the observed signal, K is the number of bases of $\boldsymbol{F}$, and L is the number of bases of $\boldsymbol{H}$. In SNMF, the matrices $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{U}$ are optimized under the condition that $\boldsymbol{F}$ is known in advance. The matrix $\boldsymbol{F}$ can be trained by solving $\hat{\boldsymbol{Y}} = \boldsymbol{F}\hat{\boldsymbol{G}}$, where $\hat{\boldsymbol{Y}}$ is a training data spectrogram and $\hat{\boldsymbol{G}}$ is the corresponding activation matrix. Hence, $\boldsymbol{F}\boldsymbol{G}$ ideally represents the target instrumental components and $\boldsymbol{H}\boldsymbol{U}$ represents other components different from the target sounds after the decomposition.

SNMF can extract the target signal, particularly in the case of a small number of sources. However, in the case of a mixture consisting of many sources, the source extraction performance is markedly degraded because of the existence of similar-timbre instruments.

2.2 Directional Clustering and Its Hybrid Method with Superresolution-Based SNMF

Decomposition methods employing directional information for the multichannel signal have also been proposed as unsupervised techniques [7]. These methods quantize the direction via time-frequency binary masking. Such directional clustering works well, even in an underdetermined situation. However, there is an inherent problem that the sources located in the same direction cannot be separated only using directional information. Furthermore, the separated signal is likely to be distorted because the signal has many spectral chasms resulting from the binary-masking procedure as shown in Fig. 1.

To solve this problem, a hybrid method that concatenates superresolution-based SNMF after directional clustering has been proposed [4]. This SNMF algorithm explicitly utilizes index information determined by time-frequency binary masking in directional clustering. For example, if the target instrument is localized in the center cluster along with the interference, superresolution-based SNMF is only applied to the existing center components using index information (see Fig. 1). Therefore, the spectrogram of the target instrument is reconstructed using more matched bases because spectral chasms are treated as *unseen*, and these chasms have no impact on the cost function in SNMF. In addition, the components of the target instrument lost after directional clustering can be extrapolated using the supervised bases. In other words, the resolution of the target spectrogram is recovered with the superresolution by the supervised basis extrapolation. Furthermore, a regularization term is added

* "超解像型教師あり非負値行列因子分解における最適なダイバージェンスのダイバーシチ," by 北村大地, 猿渡洋, 中村哲 (奈良先端大), 近藤多伸, 高橋祐 (ヤマハ株式会社), 亀岡弘和 (東京大).

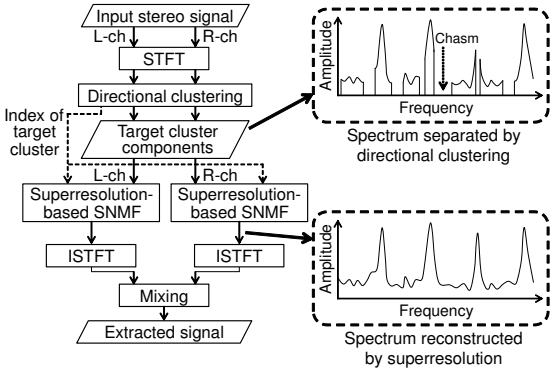

Fig. 1 Signal flow of conventional hybrid method.

in the cost function to avoid a basis extrapolation error [8].

2.3 Cost Function for Superresolution-Based SNMF

Here, the index matrix $\boldsymbol{I}(\in \mathbb{R}^{\Omega\times T}_{\{0,1\}})$ is obtained from the binary masking preceding the directional clustering. This index matrix has specific entries of unity or zero, which indicate whether or not each grid of the spectrogram belongs to the target directional cluster. The cost function in superresolution-based SNMF is defined using the index matrix $\boldsymbol{I}$ as [5]

$$\begin{aligned}\mathcal{J} = &\textstyle\sum_{\omega,t} i_{\omega,t}\mathcal{D}_{\beta_{\mathrm{NMF}}}\left(y_{\omega,t}\|\sum_k f_{\omega,k}g_{k,t} + \sum_l h_{\omega,l}u_{l,t}\right)\\ &+ \lambda\textstyle\sum_{\omega,t}\overline{i_{\omega,t}}\mathcal{D}_{\beta_{\mathrm{reg}}}(0\|\sum_k f_{\omega,k}g_{k,t}) + \mu\|\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\|^2_{\mathrm{Fr}}, \quad (2)\end{aligned}$$

where $i_{\omega,t}$, $y_{\omega,t}$, $f_{\omega,k}$, $g_{k,t}$, $h_{\omega,l}$, and $u_{l,t}$ are the nonnegative entries of matrices $\boldsymbol{I}$, $\boldsymbol{Y}$, $\boldsymbol{F}$, $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{U}$, respectively, λ and μ are the weighting parameters for each penalty term, $\bar{\cdot}$ represents the binary complement of each entry in the index matrix, and $\circ$ indicates the Hadamard product of matrices. In addition, $\mathcal{D}_\beta(\cdot\|\cdot)$ is β-divergence, which is defined as [9]

$$\mathcal{D}_\beta(y\|x) = \begin{cases} \dfrac{y^\beta}{\beta(\beta-1)} + \dfrac{x^\beta}{\beta} - \dfrac{yx^{\beta-1}}{(\beta-1)} & (\beta\in\mathbb{R}_{\setminus\{0,1\}}) \\ y(\log y - \log x) + x - y & (\beta=1) \\ \dfrac{y}{x} - \log\dfrac{y}{x} - 1 & (\beta=0) \end{cases}. \quad (3)$$

This generalized divergence is a family of cost functions parameterized by a single shape parameter β that takes IS-divergence, KL-divergence, and EUC-distance in special cases (β = 0, 1, and 2, respectively).

3 Poposed Method

3.1 Divergence Dependency on Local Chasm Condition

In general SNMF-based music signal separation, KL-divergence is often used as a cost function because KL-divergence-based SNMF decomposes the observed spectrogram into the mixture of more sparser bases compared with EUC-distance-based SNMF. The spectrogram of the music signals tends to become sparse, and KL-divergence-based SNMF fits to represent such sparse signals [6]. However, it has been experimentally confirmed that the optimal divergence for superresolution-based SNMF is EUC-distance [5]. This discrepancy in the divergence is due to the fact that superresolution-based SNMF has two tasks, namely, *signal separation* and *basis extrapolation*. The sparseness is not suitable for the basis extrapolation because it is difficult to extrapolate sparse bases only from the observable data. Figure 2 shows the dependency of the sparseness and separation performance of superresolution-based SNMF. The sparse criterion, such as KL-divergence ($\beta_{\mathrm{NMF}} = 1$), is not suitable for superresolution-based SNMF.

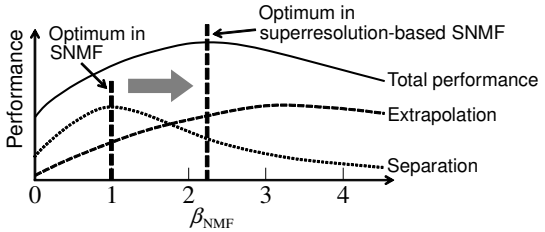

Fig. 2 Trade-off between separation and extrapolation abilities.

The optimal divergence for superresolution-based SNMF depends on the rate of spectral chasms in each time frame of the spectrogram obtained by directional clustering because of the trade-off between separation and extrapolation abilities. If there are many chasms in a frame of the binary-masked spectrogram, superresolution-based SNMF is preferred to have high extrapolation ability. In contrast, if the rate of chasms is low value, the separation ability is required rather than the extrapolation. Therefore, it is expected that EUC-distance should be used in the frames that have many chasms, and KL-divergence should be used in the other frames. To improve separation performance of superresolution-based SNMF for any types of input signals, we propose a new frame-wise divergence switching method as described below.

3.2 Cost Function

Considering the above-mentioned divergence dependency on the local chasm condition, we propose to switch the divergence in each frame of the spectrogram according to the rate of chasms in each frame, r_t, and a threshold value τ ($0 \le \tau \le 1$), where the rate of chasms r_t can be calculated from the index matrix $\boldsymbol{I}$. Figure 3 depicts an algorithm of the frame-wise divergence diversity. This divergence switching method is implemented by switching the cost function in each frame, as

$$\mathcal{J} = \textstyle\sum_t \mathcal{J}_t, \quad (4)$$

$$\mathcal{J}_t = \begin{cases} \sum_\omega i_{\omega,t}\mathcal{D}_{\beta=2}(y_{\omega,t}\|s^{(\mathrm{EUC})}_{\omega,t}) \\ \quad + \lambda\sum_\omega \overline{i_{\omega,t}}\mathcal{D}_{\beta_{\mathrm{reg}}}(0\|\sum_k f^{(\mathrm{EUC})}_{\omega,k}g_{k,t}) \\ \quad + \mu\|\boldsymbol{F}^{(\mathrm{EUC})\mathrm{T}}\boldsymbol{H}\|^2_{\mathrm{Fr}} & (r_t \ge \tau) \\ \sum_\omega i_{\omega,t}\mathcal{D}_{\beta=1}(y_{\omega,t}\|s^{(\mathrm{KL})}_{\omega,t}) \\ \quad + \lambda\sum_\omega \overline{i_{\omega,t}}\mathcal{D}_{\beta_{\mathrm{reg}}}(0\|\sum_k f^{(\mathrm{KL})}_{\omega,k}g_{k,t}) \\ \quad + \mu\|\boldsymbol{F}^{(\mathrm{KL})\mathrm{T}}\boldsymbol{H}\|^2_{\mathrm{Fr}} & (r_t < \tau) \end{cases}, \quad (5)$$

$$s^{(*)}_{\omega,t} = \textstyle\sum_k f^{(*)}_{\omega,k}g_{k,t} + \sum_n h_{\omega,n}u_{n,t}, \quad (6)$$

$$r_t = (\textstyle\sum_\omega \overline{i_{\omega,t}})/\Omega, \quad (7)$$

where $\boldsymbol{F}^{(\mathrm{KL})}(\in \mathbb{R}^{\Omega\times K}_{\ge 0})$ and $\boldsymbol{F}^{(\mathrm{EUC})}(\in \mathbb{R}^{\Omega\times K}_{\ge 0})$ are the supervised basis matrices trained in advance using KL-divergence and EUC-distance, respectively. Also, $f^{(\mathrm{KL})}_{\omega,k}$ and $f^{(\mathrm{EUC})}_{\omega,k}$ are the entries of $\boldsymbol{F}^{(\mathrm{KL})}$ and $\boldsymbol{F}^{(\mathrm{EUC})}$, respectively, and $* = \{\mathrm{KL}, \mathrm{EUC}\}$. The divergence is determined depending on r_t and τ in each frame. Therefore, this method can be considered as a frame-wise *diversity* of the divergence to achieve both of optimal separation and extrapolation.

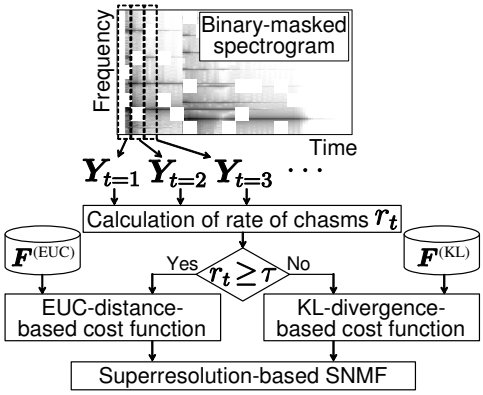

Fig. 3 Divergence diversity algorithm of proposed method.

3.3 Auxiliary Function and Update Rules

The update rules based on (4) is obtained by the auxiliary function approach. Similarly to [4, 5], we can design the upper bound function $\mathcal{J}^+$ using auxiliary variables $\alpha^{(*)}_{k,l,\omega} \geq 0$, $\gamma^{(*)}_{\omega,t,k} \geq 0$, $\delta_{\omega,t,l} \geq 0$, $\varepsilon_1 \geq 0$, $\varepsilon_2 \geq 0$, and $\zeta^{(*)}_{\omega,t} \geq 0$ that satisfy $\sum_\omega \alpha^{(*)}_{k,l,\omega} = 1$, $\sum_k \gamma_{\omega,t,k} = 1$, $\sum_l \delta_{\omega,t,l} = 1$, and $\varepsilon_1 + \varepsilon_2 = 1$, as

$$\mathcal{J} \leq \mathcal{J}^+ = \sum_t \mathcal{J}^+_t, \tag{8}$$

$$\mathcal{J}_t \leq \mathcal{J}^+_t = \begin{cases} \sum_\omega i_{\omega,t}\left(y^2_{\omega,t} + v_{\omega,t} + 2w_{\omega,t}\right) \\ \quad + \lambda \sum_\omega \overline{i_{\omega,t}} \mathcal{R}^{(\mathrm{EUC})}_{\beta_{\mathrm{reg}}} \\ \quad + \mu \sum_{k,l,\omega} \left(f^{(\mathrm{EUC})2}_{\omega,k} h^2_{\omega,l}\right) / \alpha^{(\mathrm{EUC})}_{k,l,\omega} & (r_t \geq \tau) \\ \sum_\omega i_{\omega,t}\left(-y_{\omega,t} \sum_{k,l} \gamma^{(\mathrm{KL})}_{\omega,t,k} \delta_{\omega,t,l} Q + C\right) \\ \quad + \lambda \sum_\omega \overline{i_{\omega,t}} \mathcal{R}^{(\mathrm{KL})}_{\beta_{\mathrm{reg}}} \\ \quad + \mu \sum_{k,l,\omega} \left(f^{(\mathrm{KL})2}_{\omega,k} h^2_{\omega,l}\right) / \alpha^{(\mathrm{KL})}_{k,l,\omega} & (r_t < \tau) \end{cases}, \tag{9}$$

where

$$v_{\omega,t} = \sum_k (f^{(\mathrm{EUC})2}_{\omega,k} g^2_{k,t}) / \gamma^{(\mathrm{EUC})}_{\omega,t,k} + \sum_l (h_{\omega,l} u_{l,t}) / \delta_{\omega,t,l}, \tag{10}$$

$$w_{\omega,t} = (\sum_k f^{(\mathrm{EUC})}_{\omega,k} g_{k,t})(\sum_l h_{\omega,l} u_{l,t}), - y_{\omega,t} \sum_k f^{(\mathrm{EUC})}_{\omega,k} g_{k,t} - y_{\omega,t} \sum_l h_{\omega,l} u_{l,t}, \tag{11}$$

$$\mathcal{R}^{(*)}_{\beta_{\mathrm{reg}}} = \begin{cases} \zeta^{(*)\beta_{\mathrm{reg}}-1}_{\omega,t} (\sum_k f^{(*)}_{\omega,k} g_{k,t} - \zeta^{(*)}_{\omega,t}) + \zeta^{(*)\beta_{\mathrm{reg}}}_{\omega,t} / \beta_{\mathrm{reg}} & (\beta_{\mathrm{reg}} < 1) \\ \sum_k \gamma_{\omega,t,k} (f^{(*)}_{\omega,k} g_{k,t} / \gamma^{(*)}_{\omega,t,k})^{\beta_{\mathrm{reg}}} / \beta_{\mathrm{reg}} & (1 \leq \beta_{\mathrm{reg}}) \end{cases}, \tag{12}$$

$$Q = \varepsilon_1 \log \Phi + \varepsilon_2 \log \Psi, \tag{13}$$

$$\Phi = \delta_{\omega,t,l} f^{(\mathrm{KL})}_{\omega,k} g_{k,t}, \tag{14}$$

$$\Psi = \gamma^{(\mathrm{KL})}_{\omega,t,k} h_{\omega,l} u_{l,t}. \tag{15}$$

The equality in (9) holds if and only if the auxiliary variables are set as follows:

$$\alpha^{(*)}_{k,l,\omega} = (f^{(*)}_{\omega,k} h_{\omega,l}) / (\sum_{\omega'} f^{(*)}_{\omega',k} h_{\omega',l}), \tag{16}$$

$$\gamma^{(*)}_{\omega,t,k} = (f^{(*)}_{\omega,k} g_{k,t}) / (\sum_{k'} f^{(*)}_{\omega,k'} g_{k',t}), \tag{17}$$

$$\delta_{\omega,t,l} = (h_{\omega,l} u_{l,t}) / (\sum_{l'} h_{\omega,l'} u_{l',t}), \tag{18}$$

$$\varepsilon_1 = \Phi / (\Phi + \Psi), \tag{19}$$

$$\varepsilon_2 = \Psi / (\Phi + \Psi), \tag{20}$$

$$\zeta^{(*)}_{\omega,t} = \sum_k f^{(*)}_{\omega,k} g_{k,t}. \tag{21}$$

The update rules are obtained from the derivative of the upper bound function (8) w.r.t. each objective variable and substitution of the equality condition (16)–(21), as

$$g_{k,t} \leftarrow \begin{cases} g_{k,t} \cdot \dfrac{\sum_\omega i_{\omega,t} y_{\omega,t} f^{(\mathrm{EUC})}_{\omega,k}}{\sum_\omega i_{\omega,t} f^{(\mathrm{EUC})}_{\omega,k} s^{(\mathrm{EUC})}_{\omega,t} + \lambda \sum_\omega \overline{i_{\omega,t}} f^{(\mathrm{EUC})}_{\omega,k} \left(\sum_{k'} f^{(\mathrm{EUC})}_{\omega,k'} g_{k',t}\right)^{\beta_{\mathrm{reg}}}} & (r_t \geq \tau) \\ g_{k,t} \cdot \dfrac{\sum_\omega i_{\omega,t} y_{\omega,t} f^{(\mathrm{KL})}_{\omega,k} s^{(\mathrm{KL})-1}_{\omega,t}}{\sum_\omega i_{\omega,t} f^{(\mathrm{KL})}_{\omega,k} + \lambda \sum_\omega \overline{i_{\omega,t}} f^{(\mathrm{KL})}_{\omega,k} \left(\sum_{k'} f^{(\mathrm{KL})}_{\omega,k'} g_{k',t}\right)^{\beta_{\mathrm{reg}}}} & (r_t < \tau) \end{cases}, \tag{22}$$

$$h_{\omega,l} \leftarrow h_{\omega,l} \cdot \frac{\sum_t i_{\omega,t} y_{\omega,t} u_{l,t} N_{\omega,t}}{\sum_t i_{\omega,t} u_{l,t} D_{\omega,t} + \mu P_{\omega,l}}, \tag{23}$$

$$u_{l,t} \leftarrow \begin{cases} u_{l,t} \cdot \dfrac{\sum_\omega i_{\omega,t} y_{\omega,t} h_{\omega,l}}{\sum_\omega i_{\omega,t} h_{\omega,l} s^{(\mathrm{EUC})}_{\omega,t}} & (r_t \geq \tau) \\ u_{l,t} \cdot \dfrac{\sum_\omega i_{\omega,t} y_{\omega,t} h_{\omega,l} s^{(\mathrm{EUC})-1}_{\omega,t}}{\sum_\omega i_{\omega,t} h_{\omega,l}} & (r_t < \tau) \end{cases}, \tag{24}$$

where $N_{\omega,t}$, $D_{\omega,t}$, and $P_{\omega,l}$ are given by

$$N_{\omega,t} = \begin{cases} 1 & (r_t \geq \tau) \\ s^{(\mathrm{KL})-1}_{\omega,t} & (r_t < \tau) \end{cases}, \tag{25}$$

$$D_{\omega,t} = \begin{cases} s^{(\mathrm{EUC})}_{\omega,t} & (r_t \geq \tau) \\ 1 & (r_t < \tau) \end{cases}, \tag{26}$$

$$P_{\omega,l} = \begin{cases} \sum_k f^{(\mathrm{EUC})}_{\omega,k} \sum_{\omega'} f^{(\mathrm{EUC})}_{\omega',k} h_{\omega',l} & (r_t \geq \tau) \\ \sum_k f^{(\mathrm{KL})}_{\omega,k} \sum_{\omega'} f^{(\mathrm{KL})}_{\omega',k} h_{\omega',l} & (r_t < \tau) \end{cases}. \tag{27}$$

In total, the update rules of superresolution-based SNMF with frame-wise divergence diversity are defined as (22)–(24).

Table 1 Compositions of musical instruments

Composition	Melody 1	Melody 2	Midrange	Bass
C1	Oboe	Flute	Piano	Trombone
C2	Trumpet	Violin	Harpsichord	Fagotto
C3	Clarinet	Horn	Piano	Cello

Table 2 Spatial conditions of each dataset

Spatial pattern	Measure			
	1st	2nd	3rd	4th
SP1	$\theta = 45°$	$\theta = 0°$	$\theta = 0°$	$\theta = 0°$
SP2	$\theta = 45°$	$\theta = 0°$	$\theta = 0°$	$\theta = 0°$
SP3	$\theta = 45°$	$\theta = 45°$	$\theta = 45°$	$\theta = 0°$
SP4	$\theta = 45°$	$\theta = 45°$	$\theta = 45°$	$\theta = 45°$

4 Experiments

4.1 Experimental Conditions

To confirm the effectiveness of the proposed algorithm, we compared five methods, namely, penalized SNMF (PSNMF) [2] based on KL-divergence, PSNMF based on EUC-distance, the conventional hybrid method using only EUC-distance, the conventional hybrid method using only KL-divergence, and the proposed hybrid method that switches the divergence to the optimal one framewisely. In this experiment, we used stereo signals containing four melody parts (depicted in Fig. 4) with three compositions (C1–C3) of instruments shown in Table 1. These signals were artificially generated by a MIDI synthesizer, and the observed signals $\boldsymbol{Y}$ were produced by mixing four sources with the same power. The sources were mixed as Fig. 5, where the target source was always located in the center direction with another interfering source. However, these stereo signals were mixed down to a monaural format only when we evaluate the separation accuracy of PSNMF because PSNMF is a separation method for a monaural input signal.

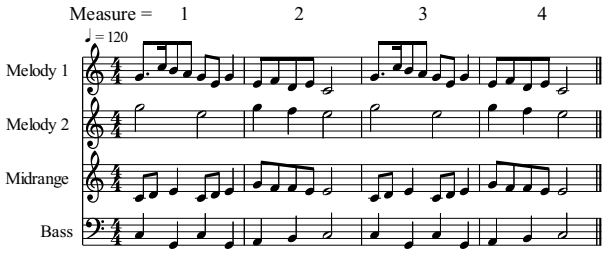

Fig. 4 Scores of each part. The observed signal consists of four measures.

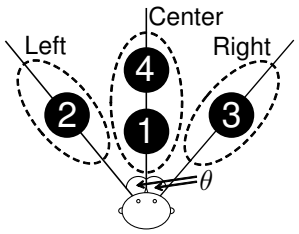

Fig. 5 Panning of four sources, where numbered black circles represent locations of instruments.

We prepared four spatially different dataset patterns of the observed signals, SP1–SP4, as shown in Table 2. In the hybrid method, many chasms were produced by directional clustering in the measures where $\theta = 45°$ compared with those of $\theta = 0°$. Therefore, we can expect that EUC-distance-based hybrid method is suitable for SP4 rather than the dataset of SP1.

In addition, we used the same MIDI sounds of the target instruments as supervision for a priori training. The training sounds contained two octave notes that cover all notes of the target signal in the observed signal. The sampling frequency of all signals was 44.1 kHz. The spectrograms were computed using a 92-ms-long rectangular window with a 46-ms overlap shift. The number of iterations for training and separation were 500, and the threshold value τ was set to 20%. Moreover, the number of clusters used in directional clustering was 3, the number of a priori bases was 100, and the number of bases for matrix $\boldsymbol{H}$ was 30. The weighting parameters λ and μ were empirically determined.

4.2 Experimental Results

We used the signal-to-distortion ratio (SDR) defined in [10] as the evaluation scores. SDR indicates the quality of the separated target sound, which includes the degree of separation between the target and other sounds and the absence of artificial distortion.

Figure 6 shows the average SDR scores for each method and each dataset pattern, where four instruments are shuffled with 12 combinations in each of compositions C1–C3. Therefore, these results are the averages of 36 input signals. From this result, KL-divergence-based hybrid method achieves high separation accuracy for the dataset of spatial patterns SP1, SP2, and SP3 because these signals do not have much spectral chasms. On the other hand, EUC-divergence-based hybrid method achieves high separation accuracy for SP 4. This dataset has many spectral chasms because the signals are always mixed with a wide panning ($\theta = 45°$), which yields many chasms, and the extrapolation ability is highly required. In addition, the proposed hybrid method with frame-wise divergence diversity can always achieve better separation for any datasets regardless of the condition whether many chasms exist or not. This is because the proposed method provides the appropriate diversity of

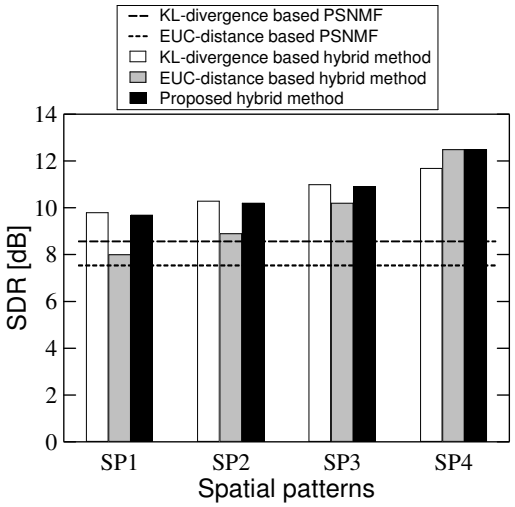

Fig. 6 Average SDR scores of each method and each spatial condition.

the divergence and can automatically apply the optimal divergence to each time frame.

5 Conclusion

In this paper, we propose a new divergence selection method to separate the target signal using optimal divergence. The proposed method switches the optimal divergence in each time frame using a threshold value for the rate of the chasms to separate and extrapolate the target signal with high accuracy. Experimental results show that our proposed method can always achieve high separation performance under any spatial conditions.

References

[1] D. D. Lee, H. S. Seung, "Algorithms for non-negative matrix factorization," *Proc. Advances in Neural Information Processing Systems*, vol.13, pp.556–562, 2001.

[2] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, K. Kondo, "Robust Music Signal Separation Based on Supervised Nonnegative Matrix Factorization with Prevention of Basis Sharing," *Proc. ISSPIT 2013*, 2013.

[3] A. Ozerov, C. Fevotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech and Language Processing*, vol.18, no.3, pp.550–563, 2010.

[4] D. Kitamura, H. Saruwatari, K. Shikano, K. Kondo, Y. Takahashi, "Superresolution-based stereo signal separation via supervised nonnegative matrix factorization," *Proc. DSP*, no.T3C-2, 2013.

[5] D. Kitamura, H. Saruwatari, K. Shikano, K. Kondo, Y. Takahashi, H. Kameoka, "Divergence optimization based on trade-off between separation and extrapolation abilities in superresolution-based nonnegative matrix factorization," *Proc. 2013 Autumn Meeting of ASJ*, 1-1-6, pp.583–586, 2013.

[6] D. FitzGerald, M. Cranitch, E. Coyle, "On the use of the beta divergence for musical source separation," *Proc. Irish Signals and Systems Conference*, 2009.

[7] S. Araki, H. Sawada, R. Mukai, S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol.87, no.8, pp.1833–1847, 2007.

[8] D. Kitamura, H. Saruwatari, K. Shikano, K. Kondo, Y. Takahashi, "Superresolution-based stereo signal separation with regularization of supervised basis extrapolation," *Proc. 3DSA 2013*, no.S10-4, 2013.

[9] S. Eguchi, Y. Kano, "Robustifying maximum likefood estimation," *Technical Report of Institute of Statistical Mathematics*, 2001.

[10] E. Vincent, R. Gribonval, C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech and Language Processing*, vol.14, no.4, pp.1462–1469, 2006.