基底共有型非負値行列因子分解に基づく楽器音の共通・固有成分の分析 *

☆香西海斗, 北村大地 (香川高専)

1　はじめに

一般的に，個人の演奏や楽器本体の芸術的価値は，一定の品質を超える範囲において，評価者の主観に基づいて評価される．例えば,「アマチュア奏者とプロフェッショナル奏者の演奏の差異」や「安価な楽器と高価な楽器の奏でる音の違い」が主観的に語られることは多い．しかしながら，これらの観点について主観を廃して定量的に議論する方法はあまり確立されておらず，とくに音楽演奏や楽器に対して精通していない者にとっては，芸術的価値を判断する材料がないため他者の主観的評価に頼らざるを得ない．もし，複数の楽器音信号の音響的違いや芸術的違いを客観的かつ定量的に表現・評価する音響特徴量が推定できるならば，アマチュア奏者がより良い演奏をするために必要な技術の提示や，より芸術的価値の高い楽器の設計製作等に役立てることができる．さらに，そのような音響特徴量は，楽器音信号の音色変換，識別，検索等にも応用することが可能と考えられ，音楽という万人が親しむ芸術文化のさらなる興隆につながることが予想される．

音響信号の特徴量解析法は，音声解析を起源として古くから研究されており，ピッチ，スペクトル包絡，メル周波数ケプストラム係数等様々な特徴量が広く利用されている．楽器音の物理現象を対象とした解析も歴史は古く，ピアノやヴァイオリン等，個々の楽器の物理音響的側面から発音機構が解析されてきた [1]．2000 年以降では，楽器同定の分野で楽器音の音響特徴量が各種検討されている [2, 3]．

本稿では，前述の目的を達成するために，客観的かつ定量的に複数の楽器音信号の違いを議論できる音響特徴量抽出法を検討する．具体的には，非負値行列因子分解（nonnegative matrix factorization: NMF）[4, 5] と呼ばれる行列分解理論を用いて，複数の楽器音信号間の「共通する音響特徴量」及び「固有の音響特徴量」を同時に推定・抽出する新しいアルゴリズムを提案し，その有用性を実験的に調査する．

2　音響信号に対する NMF の適用

2.1　NMF の概要

NMF [4, 5] は，次式に示すように，非負の観測行列を別の二つの非負行列の行列積に近似的に分解する数理アルゴリズムである．

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H} \tag{1}$$

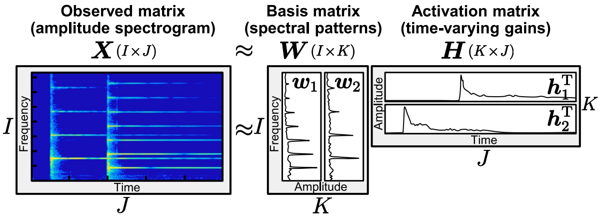

Fig. 1　NMF for audio signals, where $K = 2$.

ここで，$\boldsymbol{X} \in \mathbb{R}_{\geq 0}^{I\times J}$ は全要素が非負の観測行列であり，$\boldsymbol{W} = [\boldsymbol{w}_1\ \boldsymbol{w}_2\ \cdots\ \boldsymbol{w}_K] \in \mathbb{R}_{\geq 0}^{I\times K}$ 及び $\boldsymbol{H} = [\boldsymbol{h}_1\ \boldsymbol{h}_2\ \cdots\ \boldsymbol{h}_K]^{\mathrm{T}} \in \mathbb{R}_{\geq 0}^{K\times J}$ は NMF で推定すべき非負変数行列である．また，$\cdot^{\mathrm{T}}$ は転置を表す．$\boldsymbol{W}$ 及び $\boldsymbol{H}$ はそれぞれ基底行列及びアクティベーション行列と呼ばれる．$\boldsymbol{w}_k$ は基底ベクトルと呼ばれ，その本数 K は $K \ll \min(I, J)$ となるように設定される．ここで，$k = 1, 2, \cdots, K$ は基底ベクトルのインデクスを示す．従って，NMF は $\boldsymbol{X}$ を $\boldsymbol{W}\boldsymbol{H}$ で低ランク近似する行列分解であり，$\boldsymbol{X}$ 中に頻出する少数（K 個）の潜在的なパターンを基底ベクトルとして抽出 [4] できる教師無し学習である．

NMF を音響信号に適用する場合，短時間フーリエ変換（short-time Fourier transform: STFT）を経て得られる振幅（又はパワー）スペクトログラムを非負観測行列 $\boldsymbol{X}$ とするのが一般的である．この場合，Fig. 1 に示すように，音響信号中の頻出スペクトルが $\boldsymbol{w}_k$ として得られ，さらに各スペクトルの時間的強度変化が $\boldsymbol{h}_k$ となる．このように，NMF は音響信号中のスペクトルパターンを教師無し学習できるため，音楽信号解析 [6] や音源分離 [7, 8] に頻繁に適用される．

2.2　NMF における変数行列の最適化

NMF では，次式の最適化問題を解くことで変数行列 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ を推定する．

$$\min_{\boldsymbol{W},\boldsymbol{H}} \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) \ \text{s.t.}\ w_{ik}, h_{kj} \geq 0\ \forall i, j, k \tag{2}$$

ここで，w_{ik} 及び h_{kj} はそれぞれ $\boldsymbol{W}$ 及び $\boldsymbol{H}$ の要素であり．$i = 1, 2, \cdots, I$ 及び $j = 1, 2, \cdots, J$ はそれぞれ周波数ビン及び時間フレームのインデクスを示す．また，$\mathcal{D}(\cdot|\cdot)$ は 2 つの入力行列間の類似度を測る関数である．本稿では，次式で表される一般化 Kullback–Leibler ダイバージェンスを用いる．

$$\mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) = \sum_{i,j}\left(x_{ij}\log\frac{x_{ij}}{\sum_k w_{ik}h_{kj}} - x_{ij} + \sum_k w_{ik}h_{kj}\right) \tag{3}$$

* Analysis of common and individual components in musical instruments based on basis-shared nonnegative matrix factorization. By Kaito Kozai and Daichi Kitamura (NIT Kagawa).

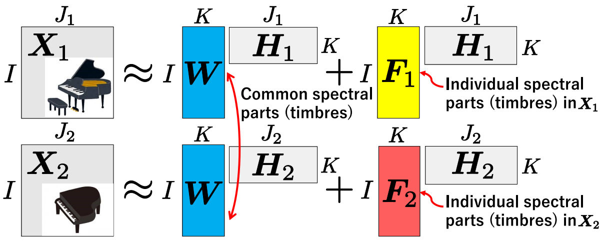

Fig. 2 Decomposition model in proposed basis-shared NMF, where $N = 2$.

3 提案手法：基底共有型 NMF

3.1 分解モデル

提案手法では，複数の音響信号間の共通成分及び固有成分を NMF で推定する．いま，N 個の観測信号の振幅スペクトログラムを $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n, \cdots, \boldsymbol{X}_N \in \mathbb{R}_{\geq 0}^{I \times J_n}$（$n = 1, 2, \cdots, N$ は観測信号のインデクス）と表すとき，次の基底共有型 NMF モデルを考える．

$$\begin{cases} \boldsymbol{X}_1 \approx \boldsymbol{W}\boldsymbol{H}_1 + \boldsymbol{F}_1\boldsymbol{H}_1 \\ \boldsymbol{X}_2 \approx \boldsymbol{W}\boldsymbol{H}_2 + \boldsymbol{F}_2\boldsymbol{H}_2 \\ \quad \vdots \\ \boldsymbol{X}_N \approx \boldsymbol{W}\boldsymbol{H}_N + \boldsymbol{F}_N\boldsymbol{H}_N \end{cases} \tag{4}$$

ここで，$\boldsymbol{W}$ は全観測信号のモデルで共有される基底行列であり，$\boldsymbol{X}_1, \cdots, \boldsymbol{X}_N$ 間の共通スペクトル成分を含む．また，$\boldsymbol{X}_n$ における固有スペクトル成分は，基底行列 $\boldsymbol{F}_n = [\boldsymbol{f}_{n1}\ \boldsymbol{f}_{n2}\ \cdots\ \boldsymbol{f}_{nK}] \in \mathbb{R}_{\geq 0}^{I \times K}$ 及びアクティベーション行列 $\boldsymbol{H}_n = [\boldsymbol{h}_{n1}\ \boldsymbol{h}_{n2}\ \cdots\ \boldsymbol{h}_{nK}]^{\mathrm{T}} \in \mathbb{R}_{\geq 0}^{K \times J_n}$ の積 $\boldsymbol{F}_n\boldsymbol{H}_n$ として表現されるため，式 (4) の分解によって N 個の観測信号中の共通及び固有成分がそれぞれ推定できる．このとき，固有成分のアクティベーション行列 $\boldsymbol{H}_n$ を共有基底行列 $\boldsymbol{W}$ と固有基底行列 $\boldsymbol{F}_n$ の間で共有することで，K 本の基底ベクトル（$\boldsymbol{w}_k$ 及び $\boldsymbol{f}_k$）のそれぞれに対する共通・固有成分への分解を実現している．

3.2 反復更新式の導出

式 (4) の分解モデルの各変数は，次の最適化問題を解くことで推定できる．

$$\min_{\boldsymbol{W}, \boldsymbol{F}_1, \cdots, \boldsymbol{F}_N, \boldsymbol{H}_1, \cdots, \boldsymbol{H}_N} \sum_n \mathcal{D}\left(\boldsymbol{X}_n | \boldsymbol{W}\boldsymbol{H}_n + \boldsymbol{F}_n\boldsymbol{H}_n\right)$$
$$\text{s.t. } w_{ik}, f_{ikn}, h_{kj_nn} \geq 0\ \forall i, j_n, k, n \tag{5}$$

ここで，$j_n = 1, 2, \cdots, J_n$ は $\boldsymbol{X}_n$ の時間フレームのインデクスを示す．以後，式 (5) の最適化を解くための反復更新式を導出する．この反復更新式の導出には，通常の NMF と同様に補助関数法 [9] を用いる．

式 (5) は次のように変形することができる．

$$\mathcal{J} \equiv \sum_n \mathcal{D}(\boldsymbol{X}_n | \boldsymbol{W}\boldsymbol{H}_n + \boldsymbol{F}_n\boldsymbol{H}_n)$$
$$= \sum_{n,i,j_n} \Bigg[x_{ij_nn} \log x_{ij_nn} - x_{ij_nn} \log\left(\sum_k w_{ik}h_{kj_nn} + \sum_k f_{ikn}h_{kj_nn} \right) - x_{ij_nn} + \sum_k w_{ik}h_{kj_nn} + \sum_k f_{ikn}h_{kj_nn} \Bigg] \tag{6}$$

上式の第 2 項について，Jensen の不等式を適用する．

$$-\log\left(\sum_k w_{ik}h_{kj_nn} + \sum_k f_{ikn}h_{kj_nn} \right)$$
$$\leq -\sum_k \alpha_{ij_nkn} \log \frac{w_{ik}h_{kj_nn}}{\alpha_{ij_nkn}} - \sum_k \beta_{ij_nkn} \log \frac{f_{ikn}h_{kj_nn}}{\beta_{ij_nkn}} \tag{7}$$

ここで，$\alpha_{ij_nkn} > 0$ 及び $\beta_{ij_nkn} > 0$ は $\sum_k \alpha_{ij_nkn} + \sum_k \beta_{ij_nkn} = 1$ を満たす補助変数である．式 (7) の等号条件を次に示す．

$$\alpha_{ij_nkn} = \frac{w_{ik}h_{kj_nn}}{\sum_{k'} w_{ik'}h_{k'j_nn} + \sum_{k'} f_{ik'n}h_{k'j_nn}} \tag{8}$$

$$\beta_{ij_nkn} = \frac{f_{ikn}h_{kj_nn}}{\sum_{k'} w_{ik'}h_{k'j_nn} + \sum_{k'} f_{ik'n}h_{k'j_nn}} \tag{9}$$

従って，$\mathcal{J} \leq \mathcal{J}^+$ なる補助関数 $\mathcal{J}^+$ が次式として設計できる．

$$\mathcal{J} \leq \mathcal{J}^+ \equiv \sum_{n,i,j_n} \Bigg(x_{ij_nn} \log x_{ij_nn} - x_{ij_nn} \sum_k \alpha_{ij_nkn} \log \frac{w_{ikn}h_{kj_nn}}{\alpha_{ij_nkn}} - x_{ij_nn} \sum_k \beta_{ij_nkn} \log \frac{f_{ikn}h_{kj_nn}}{\beta_{ij_nkn}} - x_{ij_nn} + \sum_k w_{ik}h_{kj_nn} + \sum_k f_{ikn}h_{kj_nn} \Bigg)$$

$\partial\mathcal{J}^+/\partial w_{ik} = 0$ より，$\boldsymbol{W}$ の更新式が得られる．

$$\sum_{n,j_n} \left(-\frac{x_{ij_nn}\alpha_{ij_nkn}}{w_{ik}} + h_{kj_nn} \right) = 0$$

$$w_{ik} = \frac{\sum_{n,j_n} x_{ij_nn}\alpha_{ij_nkn}}{\sum_{n,j_n} h_{kj_nn}}$$

上式に式 (8) を代入すると，反復更新式が得られる．

$$w_{ik} \leftarrow w_{ik} \frac{\sum_{n,j_n} h_{kj_nn} \frac{x_{ij_nn}}{\sum_{k'} w_{ik'}h_{k'j_nn} + \sum_{k'} f_{ik'n}h_{k'j_nn}}}{\sum_{n,j_n} h_{kj_nn}}$$

同様に，$\partial\mathcal{J}^+/\partial f_{ikn} = 0$ 及び $\partial\mathcal{J}^+/\partial h_{kj_nn} = 0$ より，$\boldsymbol{F}$ 及び $\boldsymbol{H}_n$ の更新式はそれぞれ次式となる．

$$f_{ikn} \leftarrow f_{ikn} \frac{\sum_{j_n} h_{kj_nn} \frac{x_{ij_nn}}{\sum_{k'} w_{ik'}h_{k'j_nn} + \sum_{k'} f_{ik'n}h_{k'j_nn}}}{\sum_{j_n} h_{kj_nn}}$$

$$h_{kj_nn} \leftarrow h_{kj_nn} \frac{\sum_i \frac{(w_{ik} + f_{ikn})x_{ij_nn}}{\sum_{k'} w_{ik'}h_{k'j_nn} + \sum_{k'} f_{ik'n}h_{k'j_nn}}}{\sum_i (w_{ik} + f_{ikn})}$$

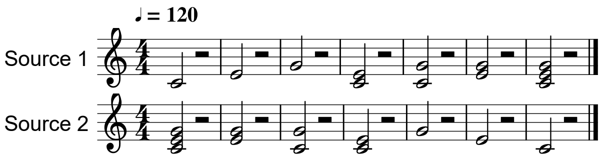

Fig. 3　Music scores used in experiment.

4　実験

4.1　実験条件

観測信号間の共通・固有成分が提案手法でどのように得られるか確認するため，MIDI 音源で作成した音響信号 $\boldsymbol{X}_1$ 及び $\boldsymbol{X}_2$ を用いて実験を行った．各音響信号は Fig. 3 に示す楽譜に基づき，ピアノ音源 Iowa Piano[1]及び 4Front Piano[2]により生成した．STFT の窓長及びシフト長はそれぞれ 92.9 ms 及び 46.4 ms とし，窓関数は Hamming 窓を用いた．提案手法の更新式の反復回数は 1000 とした．基底数は $K=6$ とし，各変数の初期値には区間 $(0,1)$ の一様乱数を用いた．

4.2　実験結果

Figs. 4(a)–4(c) はそれぞれ，共有基底行列 $\boldsymbol{W}$（$\boldsymbol{X}_1$ と $\boldsymbol{X}_2$ に共通するスペクトルパターン），$\boldsymbol{X}_1$ の固有基底行列 $\boldsymbol{F}_1$ 及び $\boldsymbol{X}_2$ の固有基底行列 $\boldsymbol{F}_2$（$\boldsymbol{X}_1$ または $\boldsymbol{X}_2$ で固有のスペクトルパターン）を示している．但し，基底行列のグラフは対数振幅の片対数グラフで示している．さらに，Figs. 5(a) 及び (b) は各スペクトルパターンのアクティベーションである $\boldsymbol{H}_1$ 及び $\boldsymbol{H}_2$ を示している．

まず，Fig. 3 と Fig. 5 を比較すると，C4 音，E4 音，及び G4 音がそれぞれ基底ベクトル 2 本ずつでモデル化されていることが分かる．また，Fig. 5 を見ると，各基底行列の偶数本目の基底ベクトル（$\boldsymbol{w}_2, \boldsymbol{w}_4, \boldsymbol{w}_6$ と $\boldsymbol{f}_{n2}, \boldsymbol{f}_{n4}, \boldsymbol{f}_{n6}$）がピアノ音のアタック部分のスペクトルを表し，奇数本目の基底ベクトル（$\boldsymbol{w}_1, \boldsymbol{w}_3, \boldsymbol{w}_5$ と $\boldsymbol{f}_{n1}, \boldsymbol{f}_{n3}, \boldsymbol{f}_{n5}$）がサスティン部分のスペクトルを表していることがわかる．確かに，Figs. 4(a) 及び (b) の $\boldsymbol{w}_1$ と $\boldsymbol{f}_{11}$ は C4 音（基本周波数 261.63 Hz）の調波構造が比較的明確に表れており，$\boldsymbol{w}_2$ と $\boldsymbol{f}_{12}$ はハンマーが弦を打つ際の非調波なスペクトルとなっている．その一方で，Fig. 4(c) の $\boldsymbol{f}_{22}$ は若干の調波構造が含まれており，この違いは $\boldsymbol{X}_1$ と $\boldsymbol{X}_2$ の 2 つのピアノ音の違いを吸収するために現れた成分（両ピアノ音のスペクトル的差異）と判断できる．同様の傾向が G4 音のアタック部分をモデル化している $\boldsymbol{f}_{16}$ 及び $\boldsymbol{f}_{26}$ にも確認できるが，E4 音のアタック部分をモデル化している $\boldsymbol{f}_{14}$ 及び $\boldsymbol{f}_{24}$ には現れなかったことから，両ピアノ音の E4 音の差異は純粋なアタック部分にのみ生じることが予測できる．

Figs. 6(a) 及び (b) は $\boldsymbol{X}_1$ 及び $\boldsymbol{X}_2$ のスペクトログラム，Figs. 6(c) 及び (d) は各音源の共通成分のスペクトログラム，Figs. 6(e) 及び (f) は各音源の固有成分のスペクトログラムをそれぞれ示している．なお，全スペクトログラムについてカラーマップは共通化している．これらを見ると，共通成分 $\boldsymbol{W}\boldsymbol{H}_1$ 及び $\boldsymbol{W}\boldsymbol{H}_2$ は比較的パワーが小さく，大部分が固有成分 $\boldsymbol{F}_1\boldsymbol{H}_1$ 及び $\boldsymbol{F}_2\boldsymbol{H}_2$ で表現されていることが分かる．もし基底数 K が十分大きければ，式 (5) の最小化は $\boldsymbol{X}_n \approx \boldsymbol{F}_n\boldsymbol{H}_n$ とモデル化することで最小化できてしまうため，固有成分 $\boldsymbol{F}_n\boldsymbol{H}_n$ のパワーが大きくなる傾向にあることは，現時点での提案手法の課題である．それでもなお，共通成分 $\boldsymbol{W}\boldsymbol{H}_n$ にはある程度の構造が抽出されており，提案手法が複数の音響信号間の共通・固有成分抽出に活用できる可能性を示している．

5　おわりに

本稿では，複数の音響信号間の共通・固有成分をモデル化するための新しい手法として，基底共有型 NMF を新たに提案した．2 種類のピアノ音を題材とした実験では，両ピアノ音のスペクトルにおける共通・固有成分を抽出することができた．今後の課題として，得られたモデルから定量的に両音響信号の違いを議論する方法の確立や，本手法を応用した音色変換等が挙げられる．

謝辞　本研究の一部は JSPS 科研費 19K20306 の助成を受けたものである．

参考文献

[1] N. H. Fletcher and T. D. Rossing, “The physics of musical instruments,” *Springer Science & Business Media*, 1991.

[2] T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, “Musical instrument recognizer “instrogram” and its application to music retrieval based on instrumentation similarity,” *Proc. Int. Symp. Multimedia*, pp. 265–274, 2006.

[3] C. Joder, S. Essid, and G. Richard, “Temporal integration for audio classification with application to musical instrument classification,” *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 17, no. 1, pp. 174–186, 2009.

[4] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[5] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization” *Proc. Neural Info. Process. Syst.*, pp. 556–562, 2000.

[6] C. Févotte, N. Bertin, and J.-L. Durrieu, “Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis,” *Neural Comput.*, vol. 21, no. 3, pp. 793–830, 2009.

[7] D. FitzGerald, M. Cranitch, and E. Coyle, “On the use of the beta divergence for musical source separation,” *Proc. Irish Signals Syst. Conf.*, 2009.

[8] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, “Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties,” *IEICE Trans. Fundamentals*, vol.E97-A, no.5, pp.1113–1118, 2014.

[9] D. R. Hunter and K. Lange, “A tutorial on MM algorithms,” *The American Statist.*, vol. 58, no. 1, pp. 30–37, 2004.

---

[1] `https://vst4free.com/plugin/2294/`
[2] `http://www.yohng.com/software/piano.html`

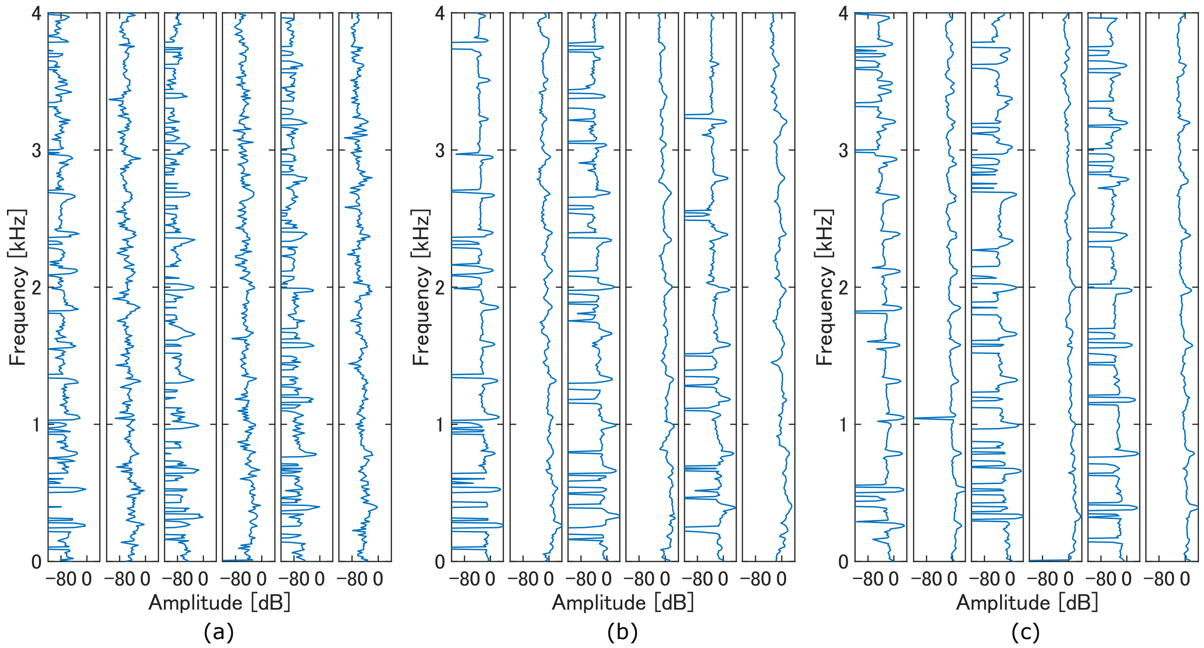

Fig. 4 Spectral patterns in basis matrices (a) $\boldsymbol{W}$, (b) $\boldsymbol{F}_1$, and (c) $\boldsymbol{F}_2$ estimated by proposed basis-shared NMF.

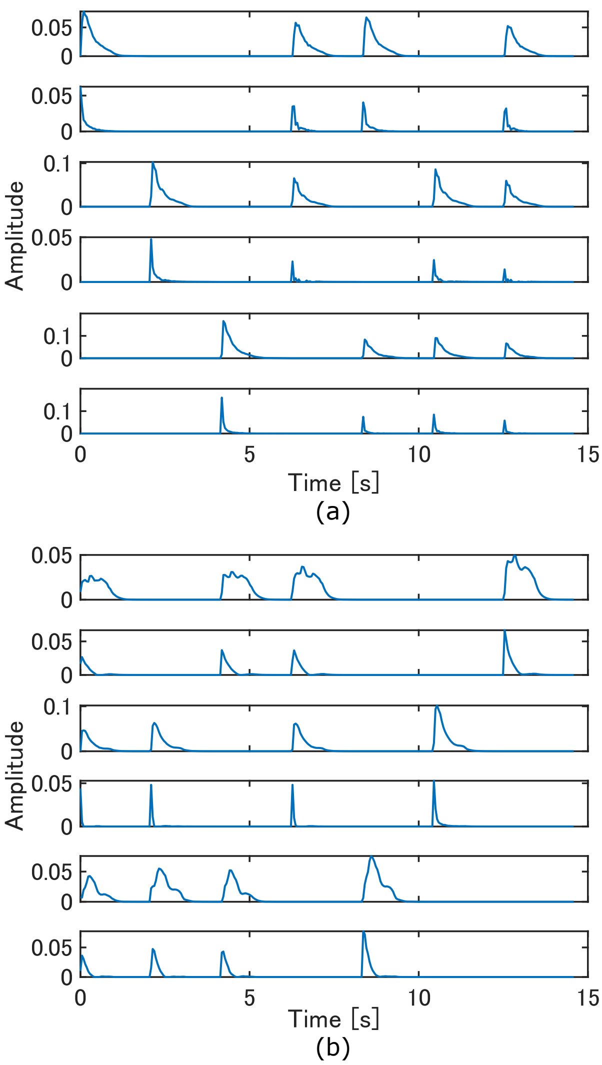

Fig. 5 Time-varying gains in activation matrices (a) $\boldsymbol{H}_1$ and (b) $\boldsymbol{H}_2$ estimated by proposed basis-shared NMF.

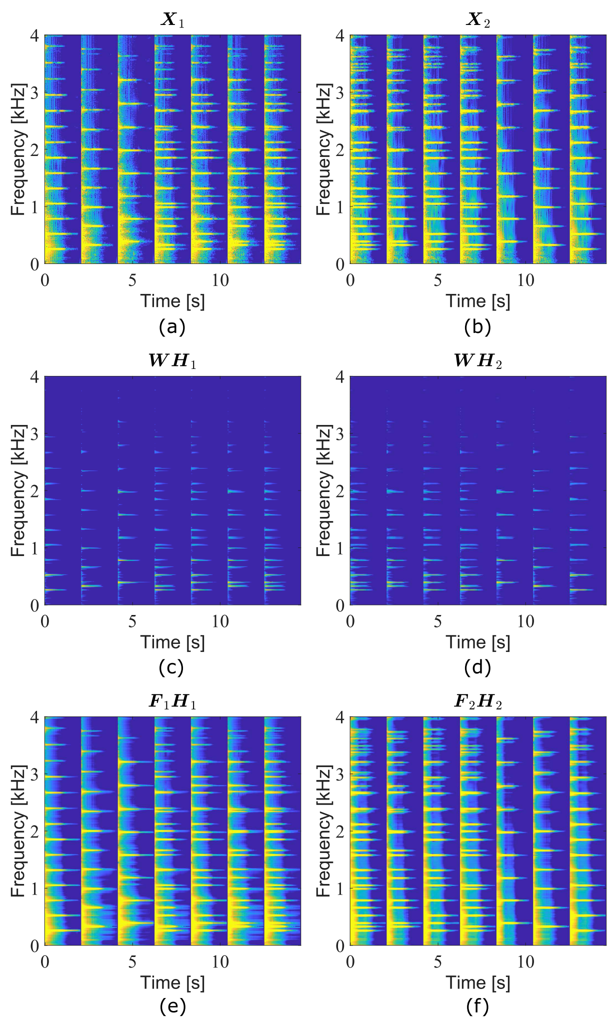

Fig. 6 Spectrograms of observed signals (a) $\boldsymbol{X}_1$ and (b) $\boldsymbol{X}_2$ and decomposed model spectrograms estimated by proposed basis-shared NMF: common components (c) $\boldsymbol{W}\boldsymbol{H}_1$ and (d) $\boldsymbol{W}\boldsymbol{H}_2$ and individual components (e) $\boldsymbol{F}_1\boldsymbol{H}_1$ and (f) $\boldsymbol{F}_2\boldsymbol{H}_2$.