メディアン型HPSSを用いた時間周波数マスクに基づくブラインド音源分離*

☆大藪宗一郎，北村大地（香川高専），矢田部浩平（早稲田大）

1 はじめに

ブラインド音源分離（blind source separation: BSS）[1] とは，マイクロホンや音源の位置等の事前情報を用いずに，複数の音源が混合した観測信号から，混合前の音源信号を推定する技術である．観測チャネル数が音源数以上となる優決定条件でのBSSには，独立成分分析（independent component analysis: ICA）[2] に基づく手法が広く用いられている．例えば，独立ベクトル分析（independent vector analysis: IVA）[3, 4] 及び独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[5, 6] 等が提案されている．これらの手法では，音源信号に関する事前知識（音源モデル）に基づいてパーミュテーション問題 [7] を解決している．このとき，音源モデルが混合前の音源信号に適合しているか否かで性能が左右される．より良い音源モデルをBSSに導入できれば，より高品質な分離信号が得られる可能性があるため，種々の音源モデルを用いたBSSで性能を比較することが重要である．

この目的に対し，幅広い音源モデルを統一的に扱えるBSSアルゴリズムとして，時間周波数マスキングに基づく優決定BSS（time-frequency-masking-based determined BSS: TFMBSS）[8] が提案された．TFMBSSは，時間周波数マスクで表される音源モデルを用いて，線形の（歪みの少ない）多チャネル音源分離が可能である．文献 [9] では，調波打撃音分離（harmonic/percussive source separation: HPSS）[10] に基づく時間周波数マスクをTFMBSSの音源モデルとして用いた手法を新たに提案し，その性能を調査した．さらに，文献 [11] では，時間周波数マスクの生成方法に改良を加え，より排他的な（他音源を抑圧するような）時間周波数マスクが得られる処理を新規に導入した．これらの手法は，HPSSに基づく音源モデルを活用していることから，調波音と打撃音の多チャネル音源分離に利用可能であり，音楽信号の解析（コード・テンポ・音階等の推定）等に応用できる．

本稿では，音源モデルをplug-and-playで変更できるTFMBSSの利点を活かし，別の調波打撃音分離手法であるメディアン型HPSS [12] に基づくTFMBSSを新たに提案する．メディアン型HPSSは，時間方向と周波数方向のそれぞれにメディアンフィルタを適用することで，調波打撃音分離を行うモノラル音源分離手法である．また，提案手法においてフィルタサイズやスムージングパラメータについて実験的に調査し考察する．そして，従来のHPSSに基づくTFMBSSと性能比較し有用性の検討を行う．

2 従来手法

2.1 定式化

音源数と観測チャネル数をそれぞれ N 及び M とし，多チャネル時間信号をSTFTして得られる時間周波数毎の音源信号，観測信号，及び分離信号をそれぞれ

$$\boldsymbol{s}_{ij} = [s_{ij1},\ \cdots, s_{ijn},\ \cdots,\ s_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \quad (1)$$

$$\boldsymbol{x}_{ij} = [x_{ij1},\ \cdots, x_{ijm},\ \cdots,\ x_{ijM}]^{\mathrm{T}} \in \mathbb{C}^M \quad (2)$$

$$\boldsymbol{y}_{ij} = [y_{ij1},\ \cdots, y_{ijn},\ \cdots,\ y_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \quad (3)$$

と表す．ここで，$i=1,2,\cdots,I$，$j=1,2,\cdots,J$，$n=1,2,\cdots,N$，及び $m=1,2,\cdots,M$ はそれぞれ周波数ビン，時間フレーム，音源，及びチャネルのインデクスを示し，$\cdot^{\mathrm{T}}$ は転置を表す．また，各信号の複素スペクトログラムを $\boldsymbol{S}_n \in \mathbb{C}^{I\times J}$，$\boldsymbol{X}_m \in \mathbb{C}^{I\times J}$，及び $\boldsymbol{Y}_n \in \mathbb{C}^{I\times J}$ で表す．

混合系が線形時不変であり，時間周波数領域での複素瞬時混合で表現できると仮定すると，観測信号と音源信号の関係を次式で表現できる．

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \quad (4)$$

ここで，$\boldsymbol{A}_i \in \mathbb{C}^{M\times N}$ は周波数毎の混合行列である．この混合モデルは，残響時間がSTFTの窓長よりも十分短い場合に近似的に成立する．以後，決定的な系（$M=N$）を考える．$\boldsymbol{A}_i$ が正則であれば，逆行列を用いて分離信号を次式で推定できる．

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \quad (5)$$

ここで，$\boldsymbol{W}_i = [\boldsymbol{w}_{i1}\ \cdots\ \boldsymbol{w}_{iN}]^{\mathrm{H}} \in \mathbb{C}^{N\times M}$ は周波数毎の分離行列であり，$\cdot^{\mathrm{H}}$ はエルミート転置である．優決定条件BSSでは，式 (5) 中の分離行列 $\boldsymbol{W}_i$ を全周波数 ($i=1,\cdots,I$) において推定することが最終的な目標となる．式 (5) で求まる分離信号 $\boldsymbol{y}_{ij}$ は，混合信号 $\boldsymbol{x}_{ij}$ に対する線形フィルタリングであり，自然性の高い音源分離が可能な利点がある．

2.2 HPSS

HPSS [10] は，振幅スペクトログラムの時間方向及び周波数方向の滑らかさに基づき，調波音及び打撃音を分離する．モノラルの混合信号，分離調波信号，分離打撃信号の複素スペクトログラムをそれぞれ $\boldsymbol{B}\in\mathbb{C}^{I\times J}$，$\boldsymbol{H}\in\mathbb{C}^{I\times J}$，及び $\boldsymbol{P}\in\mathbb{C}^{I\times J}$ と表すと，文献 [10] のHPSSでは，混合信号 $\boldsymbol{B}$ から $\boldsymbol{H}$ と $\boldsymbol{P}$ を推定するために，次式の目的関数を $\boldsymbol{H}$ 及び $\boldsymbol{P}$ に関して最小化する．

$$J(\boldsymbol{H},\boldsymbol{P}) = \sum_{i,j} \Big\{ \gamma_H \left(|h_{i(j+1)}|^{0.5} - |h_{ij}|^{0.5} \right)^2 + \gamma_P \left(|p_{(i+1)j}|^{0.5} - |p_{ij}|^{0.5} \right)^2 \Big\} \quad (6)$$

ここで，h_{ij} 及び p_{ij} はそれぞれ $\boldsymbol{H}$ 及び $\boldsymbol{P}$ の要素であり，$\gamma_H > 0$ 及び $\gamma_P > 0$ は重み係数である．式 (6) の最小化問題は，次の制約条件が課せられている．

$$|b_{ij}| = |h_{ij}| + |p_{ij}| \quad (7)$$

ここで，b_{ij} は $\boldsymbol{B}$ の要素である．式 (6) を最小化する h_{ij} 及び p_{ij} は，次の反復更新式で推定できる [10].

$$|h_{ij}|^{0.5} = \frac{\gamma_H \left(|h_{(i+1)j}|^{0.5} + |h_{(i-1)j}|^{0.5} \right) |b_{ij}|^{0.5}}{\sqrt{c_{ij}^{(H)} + c_{ij}^{(P)}}} \quad (8)$$

*Blind source separation based on time-frequency mask using median-type HPSS. By Soichiro OYABU, Daichi KITAMURA (NIT Kagawa), and Kohei YATABE (Waseda Univ.).

Algorithm 1 TFMBSS

Input: X, $\mathbf{w}^{[1]}$, $\mathbf{y}^{[1]}$, μ_1, μ_2, α
Output: $\mathbf{w}^{[k+1]}$

1: **for** $k = 1, \cdots, K$ **do**
2: $\quad \widetilde{\mathbf{w}} = \mathrm{prox}_{\mu_1 \mathcal{I}} \left[\mathbf{w}^{[k]} - \mu_1 \mu_2 X^{\mathrm{H}} \mathbf{y}^{[k]} \right]$
3: $\quad \mathbf{z} = \mathbf{y}^{[k]} + X(2\widetilde{\mathbf{w}} - \mathbf{w}^{[k]})$
4: $\quad \widetilde{\mathbf{y}} = \mathbf{z} - \mathcal{M}(\mathbf{z}) \odot \mathbf{z}$
5: $\quad \mathbf{y}^{[k+1]} = \alpha \widetilde{\mathbf{y}} + (1-\alpha)\mathbf{y}^{[k]}$
6: $\quad \mathbf{w}^{[k+1]} = \alpha \widetilde{\mathbf{w}} + (1-\alpha)\mathbf{w}^{[k]}$
7: **end for**

$$|p_{ij}|^{0.5} = \frac{\gamma_P \left(|p_{i(j+1)}|^{0.5} + |p_{i(j-1)}|^{0.5} \right) |b_{ij}|^{0.5}}{\sqrt{c_{ij}^{(H)} + c_{ij}^{(P)}}} \tag{9}$$

$$c_{ij}^{(H)} = {\gamma_H}^2 \left(|h_{(i+1)j}|^{0.5} + |h_{(i-1)j}|^{0.5} \right)^2 \tag{10}$$

$$c_{ij}^{(P)} = {\gamma_P}^2 \left(|p_{i(j+1)}|^{0.5} + |p_{i(j-1)}|^{0.5} \right)^2 \tag{11}$$

2.3 TFMBSS

TFMBSS [8] とは，時間周波数マスクで表現される音源モデルに基づく優決定条件 BSS である．TFMBSS のアルゴリズムを Algorithm 1 に示す．ここで，X は多チャネル観測信号の複素スペクトログラム（$\boldsymbol{X}_1, \cdots, \boldsymbol{X}_M$）から構成される行列，$\mathbf{w}$ は全周波数の分離行列（$\boldsymbol{W}_1, \cdots, \boldsymbol{W}_I$）をベクトル化した変数，$\odot$ は要素毎の積を表す（詳細な定義は文献 [8] 参照）．Algorithm 1 の 4 行目の $\mathcal{M}(\mathbf{z})$ が，TFMBSS で用いられる時間周波数マスクである．中間変数 $\mathbf{z}$ を引数とし分離をさらに促進するような時間周波数マスクを返す関数 $\mathcal{M}$ を音源モデルとして活用する．従って，TFMBSS では，$\mathcal{M}(\mathbf{z})$ を自由に入れ替えることで，様々な音源モデルを導入した BSS が実現される．

3 提案手法

3.1 動機

文献 [9, 11] では，調波信号と打撃信号の高品質な BSS を目的として，TFMBSS の時間周波数マスク $\mathcal{M}(\mathbf{z})$ に，式 (6) で定式化される HPSS [10] を導入した手法を提案した．この手法のブロック図を Fig. 1 に示す．本手法では，TFMBSS の最適化で得られる中間変数の $\mathbf{z}$ から調波信号と打撃信号の時間周波数成分（$\boldsymbol{Z}_H \in \mathbb{C}^{I \times J}$ 及び $\boldsymbol{Z}_P \in \mathbb{C}^{I \times J}$）を抽出し HPSS を適用することで，各信号をさらに強調する時間周波数マスク（$\mathcal{M}_H \in \mathbb{R}_{[0,1]}^{I \times J}$ 及び $\mathcal{M}_P \in \mathbb{R}_{[0,1]}^{I \times J}$）を生成している．また，TFMBSS の最適化の安定性を向上させるために，過去の時間周波数マスクとのスムージング処理を導入している．

一方で，メディアンフィルタに基づく HPSS [12] も提案されており，比較的少ない計算量で高い性能が得られることが実験的に示されている [13]．本稿では，更なる分離性能向上を目指して，Fig. 1 中の HPSS のブロックをメディアン型 HPSS に変更した BSS を提案し，従来手法 [11] と実験的に比較する．

3.2 メディアン型 HPSS

HPSS は，式 (6) の目的関数で表現される通り，スペクトログラムの時間方向及び周波数方向に滑らかな成分に分離することで調波信号と打撃信号を推定する．式 (6) の最小化とは異なる方法で，同様の原理に基づいて調波打撃音分離を達成した手法として，メディアン型 HPSS [12] が提案されている．本手法は，

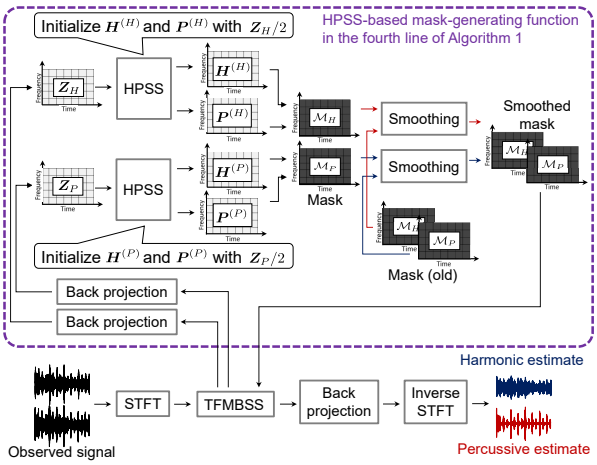

Fig. 1 Block diagram of TFMBSS based on HPSS.

スペクトログラムの時間方向及び周波数方向にそれぞれメディアンフィルタを適用する．メディアンフィルタは，フィルタを適用する方向のスパイク状の成分を除去できるため，非線形かつ強力な平滑化が施される．従って，時間方向及び周波数方向の滑らかさを強調した信号を推定することができ，調波信号及び打撃信号が得られる．

メディアン型 HPSS は，フィルタサイズ $2L+1$ のメディアンフィルタをシフト長 1 点でずらしながら適用する．メディアンフィルタを適用するベクトルは，次式のように混合信号 $\boldsymbol{B}$ の行及び列の一部となる．

$$\boldsymbol{b}_{ij}^{(\mathrm{r})} = \left[|b_{i(j-L)}|, |b_{i(j-L+1)}|, \cdots, |b_{i(j+L)}| \right] \in \mathbb{R}_{\geq 0}^{2L+1} \tag{12}$$

$$\boldsymbol{b}_{ij}^{(\mathrm{c})} = \left[|b_{(i-L)j}|, |b_{(i-L+1)j}|, \cdots, |b_{(i+L)j}| \right] \in \mathbb{R}_{\geq 0}^{2L+1} \tag{13}$$

これらのベクトルにメディアンフィルタを適用することで，h_{ij} 及び p_{ij} が推定できる．

$$|h_{ij}| = \mathrm{median}(\boldsymbol{b}_{ij}^{(\mathrm{r})}) \tag{14}$$

$$|p_{ij}| = \mathrm{median}(\boldsymbol{b}_{ij}^{(\mathrm{c})}) \tag{15}$$

ここで，$\mathrm{median}(\cdot)$ は入力されたベクトルの中央値のみをスカラーとして返す関数である．

3.3 メディアン型 HPSS に基づく TFMBSS

本稿における提案手法は，Fig. 1 の HPSS ブロックを，前節のメディアン型 HPSS に置き換えたものとなる．ここで，中間変数 $\mathbf{z}$ を，調波音成分及び打撃音成分のスペクトログラムのサイズに整形した変数をそれぞれ $\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ と表記している．従来手法 [11] のアルゴリズムを踏襲し，TFMBSS の反復最適化途中で得られる中間変数 $\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ のそれぞれに対して，独立なメディアン型 HPSS を適用する．即ち，メディアン型 HPSS による音源強調をフィルタとして考え，調波音成分 $\boldsymbol{Z}_H$ の中に残留する打撃音成分や，打撃音成分 $\boldsymbol{Z}_P$ の中に残留する調波音成分を取り除く排他的な時間周波数マスクをそれぞれ生成している．

具体的には，Fig. 1 に示すように，TFMBSS で得られる中間変数 $\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ を次式のように 2 つの

独立なメディアン型 HPSS の観測信号とする．

$$|\boldsymbol{B}^{(\boldsymbol{Z}_H)}| = |\boldsymbol{Z}_H| \tag{16}$$

$$|\boldsymbol{B}^{(\boldsymbol{Z}_P)}| = |\boldsymbol{Z}_P| \tag{17}$$

ここで，$\boldsymbol{B}^{(\boldsymbol{Z}_H)} \in \mathbb{C}^{I\times J}$ 及び $\boldsymbol{B}^{(\boldsymbol{Z}_P)} \in \mathbb{C}^{I\times J}$ は調波音成分及び打撃音成分用のメディアン型 HPSS の観測信号である．また，行列に対する演算 $|\cdot|$ は要素毎の絶対値処理を示す．$|\boldsymbol{B}^{(\boldsymbol{Z}_H)}|$ 及び $|\boldsymbol{B}^{(\boldsymbol{Z}_P)}|$ にそれぞれメディアン型 HPSS を適用するため，$\boldsymbol{Z}_H$ 中の調波音成分 $|\boldsymbol{H}^{(\boldsymbol{Z}_H)}|$ と打撃音成分 $|\boldsymbol{P}^{(\boldsymbol{Z}_H)}|$ 及び $\boldsymbol{Z}_P$ 中の調波音成分 $|\boldsymbol{H}^{(\boldsymbol{Z}_P)}|$ と打撃音成分 $|\boldsymbol{P}^{(\boldsymbol{Z}_P)}|$ の合計 4 種類の信号を推定することとなる．ここで，$\boldsymbol{Z}_H$ 中の打撃音成分 $|\boldsymbol{P}^{(\boldsymbol{Z}_H)}|$ と $\boldsymbol{Z}_P$ 中の調波音成分 $|\boldsymbol{H}^{(\boldsymbol{Z}_P)}|$ が分離途中の混合信号における残留成分に該当する．そして，得られた調波音成分と打撃音成分のパワースペクトログラムから，次式の Wiener フィルタを構築し，これを新たな時間周波数マスクとする．

$$[\mathcal{M}_H]_{ij} = \left(\frac{|h_{ij}^{(\boldsymbol{Z}_H)}|^2}{|h_{ij}^{(\boldsymbol{Z}_H)}|^2 + |p_{ij}^{(\boldsymbol{Z}_H)}|^2}\right)^{\frac{1}{2}} \tag{18}$$

$$[\mathcal{M}_P]_{ij} = \left(\frac{|p_{ij}^{(\boldsymbol{Z}_P)}|^2}{|h_{ij}^{(\boldsymbol{Z}_P)}|^2 + |p_{ij}^{(\boldsymbol{Z}_P)}|^2}\right)^{\frac{1}{2}} \tag{19}$$

ここで，$h_{ij}^{(\boldsymbol{Z}_H)}$，$p_{ij}^{(\boldsymbol{Z}_H)}$，$h_{ij}^{(\boldsymbol{Z}_P)}$，及び $p_{ij}^{(\boldsymbol{Z}_P)}$ はそれぞれ $\boldsymbol{H}^{(\boldsymbol{Z}_H)}$，$\boldsymbol{P}^{(\boldsymbol{Z}_H)}$，$\boldsymbol{H}^{(\boldsymbol{Z}_P)}$，及び $\boldsymbol{P}^{(\boldsymbol{Z}_P)}$ の要素である．さらに，$\mathcal{M}_H$ 及び $\mathcal{M}_P$ はそれぞれの観測信号中の残留成分を抑圧する時間周波数マスクであり，$[\mathcal{M}]_{ij}$ はマスク $\mathcal{M}$ の ij 要素を表す．これより，排他的な時間周波数マスクを生成し，TFMBSS を反復最適化するアルゴリズムとなっている．

TFMBSS では，時間周波数マスク $\mathcal{M}$ が反復毎に大きく変動する場合，安定した音源分離ができない場合がある．この問題に対処するために，マスクを生成する度に，次式で 1 反復前のマスク $\mathcal{M}_{\text{old}}$ とのスムージングを施す．

$$\mathcal{M} = \mathcal{M}^{\beta} \odot {\mathcal{M}_{\text{old}}}^{\beta_{\text{old}}} \tag{20}$$

ここで，β 及び β_{old} はそれぞれスムージング度合いを決定するパラメータであり，$\beta + \beta_{\text{old}} = 1$ を満たす．式 (20) の処理を $\mathcal{M}_H$ 及び $\mathcal{M}_P$ のそれぞれに施す．これを新たな時間周波数マスクとして TFMBSS に返され，Algorithm 1 の 4 行目として，中間変数 $\mathbf{z}$ 中の調波音成分と打撃音成分にそれぞれ適用される．なお，TFMBSS も IVA や ILRMA と同様に分離信号のスケールを推定できない為，TFMBSS からメディアン型 HPSS に変数を引き渡すタイミングで，中間変数 $\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ に対してプロジェクションバック法 [14] を適用し，周波数毎のスケールを補正している．

4 実験

4.1 実験条件

提案手法の有効性を確認するために，音楽信号中のドラムとそれ以外の楽器音の音源分離実験を行った．本実験では，SiSEC2016 [15] の DSD100 データセット中のドラム音源（drums）とその他の音源（other）を 20 曲選んだ．これらのドライソースを，文献 [16] に記載のマイク間隔 5.66 cm 及び音源方位 50°&130° の E2A インパルス応答 [17]（残響長 300 ms）で畳み込み，多チャネル混合信号を作成した．評価指標には，音源対歪み比（source-to-distortion ratio: SDR）[18] の改善量を用いた．その他の実験条件を Table 1 に示す．

Table 1 Experimental conditions

Window function in STFT	Hann window
Window length in STFT	128 ms
Shift length in STFT	64 ms
Parameters in TFMBSS	$\alpha = 0.25$ $\mu_1 = \mu_2 = 1.0$
Number of iterations in BSS	500

Table 2 Average SDR improvements of proposed method with various filter sizes

Filter size	Average SDR improvement [dB]
3	7.28
5	10.33
7	11.44
9	11.67
11	11.80
13	11.97
15	11.96
17	**12.04**
19	12.02
21	12.00

4.2 メディアンフィルタサイズに対する性能の変化

提案手法では，メディアンフィルタのサイズ $2L+1$ を設定してから推定を行う．このときのフィルタサイズを変化させることによる SDR 改善量の変化を調査した．

提案手法における全 20 曲の平均 SDR 改善量を Table 2 に示す．ここで，時間方向と周波数方向のフィルタサイズは同一に固定した．また，本実験におけるスムージングパラメータは $\beta_{\text{old}} = 0.75$ 及び $\beta = 0.25$ であり，これはアルゴリズムの安定性と分離性能のバランスを考慮した最適な設定値である（次節の実験結果に基づく）．Table 2 の結果より，本実験条件では，フィルタサイズは 17 点が最適であることが確認された．スムージングパラメータを変更した際，最適な数値が変動することも実験的に確認しているが，おおよそ 17 点周辺の値であった．以降の実験では，この結果に基づき，提案手法のフィルタリングサイズを 17 点と設定する．

4.3 スムージングパラメータに対する性能の変化

次に，提案手法におけるスムージングの有効性を検証する．提案手法において，β_{old} 及び β のみを変化させた場合の反復毎の SDR 改善量の一例（song no. 20）を Fig. 2 に示す．但し，常に $\beta_{\text{old}} + \beta = 1$ である．さらに，提案手法における全 20 曲の平均 SDR 改善量を Table 3 に示す．Fig. 2 の $\beta_{\text{old}} = 0$ と $\beta_{\text{old}} = 0.5$ のように低く設定すると，極端に分離性能が下がる瞬間が所々で存在する．従って，最終的な分離結果の性能が極端に下がってしまう可能性がある．一方，β_{old} を高く設定した場合（スムージング強くした場合），SDR 改善量の推移は安定するが，収束速度が遅くなるため，より多い反復最適化が要求される．また，収束時の SDR 改善量値も $\beta_{\text{old}} = 0.75$ の例と比較すると低下していることが分かる．以上より，最終的な収束値

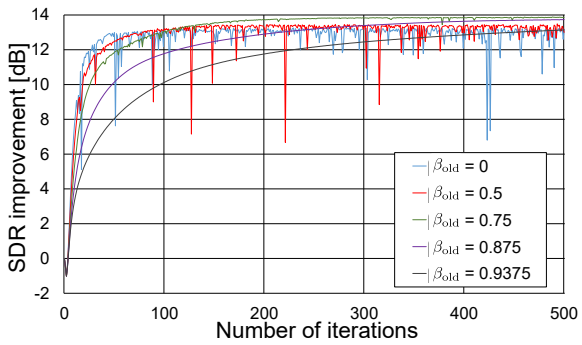

Fig. 2　Example of convergence behaviors of proposed method with various β_{old} and β (song no. 20).

Table 3　Average SDR improvements of proposed method with various smoothing parameters β and β_{old}

β	β_{old}	Average SDR improvement [dB]
1	0	11.58
0.5	0.5	11.77
0.25	0.75	**12.04**
0.125	0.875	11.73
0.0625	0.9375	10.94

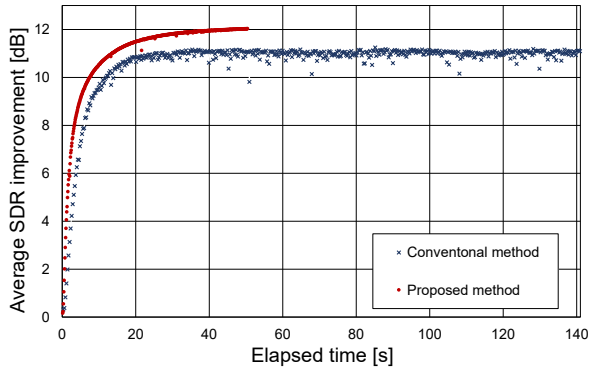

Fig. 3　Average convergence behaviors of SDR improvements in conventional and proposed methods in terms of elapsed time.

と安定性のトレードオフを考慮すると，$\beta_{\text{old}} = 0.75$ 及び $\beta = 0.25$ が最適であることが分かる．これらの傾向は，従来手法 [11] と同様であり，音源モデルには依存しないことが確認された．

4.4　従来手法との性能比較

最後に，従来手法 [11] と提案手法の性能比較を行う．但し，従来手法のパラメータは文献 [11] での実験結果における最適パラメータを用いる．これは，提案手法の最適パラメータと全く同一である．Fig. 3 は，計算に要する経過時間に対する，データセット 20 曲全ての平均 SDR 改善量の比較である．この結果より，提案手法は従来手法よりも分離性能が若干向上し，分離に要する時間は半分以下であることが確認された．よって調波音と打撃音の分離において従来手法よりも高速かつ高性能な多チャネル BSS であることが確認できる．

5　まとめ

本稿では，新たにメディアン型 HPSS の音源モデルに基づいて，マスクを生成する TFMBSS を提案した．さらに提案手法におけるパラメータの検討として，フィルタサイズ及びマスクのスムージング度合いにおいて比較実験を行い，性能の変化を確認した．提案手法は，従来の HPSS に基づく TFMBSS よりも性能改善かつ高速化を達成できることを示した．

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] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, “A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF,” *APSIPA Transactions on Signal and Information Processing*, vol. 8, no. e12, pp. 1–14, 2019.

[2] P. Comon, “Independent component analysis, a new concept?,” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994..

[3] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 70–79, 2007.

[4] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 189–192, 2011.

[5] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE/ACM Trans. Workshop on Applications of Signal Processing to Audio and Acoustics*, vol. 24, no. 9, pp. 1626–1641, 2016.

[6] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation with independent low-rank matrix analysis,” *In Audio Source Separation,* S. Makino, Ed., pp. 125–155, Springer, Cham, 2018.

[7] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *IEEE Trans. SAP*, vol. 12, no. 5, pp. 530–538, Sep. 2004.

[8] K. Yatabe and D. Kitamura, “Time-frequency-masking-based determined BSS with application to sparse IVA,” *Proc. International Conference on Acoustics, Speech, and Signal Processing*, pp. 715–719, 2019.

[9] 大藪宗一郎, 北村大地, 矢田部浩平, “調波打撃音分離の時間周波数マスクを用いた線形ブラインド音源分離,” **日本音響学会** *2020* **年春季研究発表会講演論文集**, pp. 313–316, 2020.

[10] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama, “Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram,” *Proc. European Signal Processing Conference*, 2008.

[11] 大藪宗一郎, 北村大地, 矢田部浩平, “調波打撃音分離の排他的マスキングに基づくブラインド音源分離,” **日本音響学会** *2020* **年秋季研究発表会講演論文集**, pp. 283–286, 2020.

[12] D FitzGerald, “Harmonic/percussive separation using median filtering,” *Proc. International Conference on Digital Audio Effects (DAFx)*, vol. 13, 2010.

[13] Y. Masuyama, K. Yatabe, and Y. Oikawa, “Phase-aware harmonic/percussive source separation via convex optimization,” *Proc. International Conference on Acoustics, Speech, and Signal Processing*, pp. 985–989, 2019.

[14] N. Murata, S. Ikeda, and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals,” *Neurocomputing*, vol. 41, no. 1–4, pp. 1–24, 2001.

[15] A. Liutkus, F.-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, “The 2016 signal separation evaluation campaign,” *Proc. 13th International Conference on Latent Variable Analysis and Signal Separation*, pp. 323–332, 2017.

[16] D. Kitamura, N. Ono, and H. Saruwatari, “Experimental analysis of optimal window length for independent low-rank matrix analysis,” *Proc. EUSIPCO*, pp. 1210–1214, 2017.

[17] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, “Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition,” *Proc. Language Resources and Evaluation Conference*, pp. 965–968, 2000.

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