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*

☆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）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）[1] 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）[2, 3] 及び独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[4, 5] 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 [6] 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）[7] 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 [8] では，調波打撃音分離（harmonic/percussive source separation: HPSS）[9] 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．本手法は HPSS に基づいていることから，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，この手法の時間周波数マスクの生成方法を改良し，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 TFMBSS の内部に導入する．また，従来手法と提案手法の両方において，HPSS の最適化の反復回数やマスクの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．

2 従来手法：HPSS に基づく TFMBSS

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，$\boldsymbol{W}_i = [\boldsymbol{w}_{i1}\ \cdots\ \boldsymbol{w}_{iN}]^{\mathrm{H}} \in \mathbb{C}^{N\times N}$ 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 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 [9] 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，式 (6) 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．

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

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 [9]．

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

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

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

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

*Blind source separation based on exclusive mask obtained by harmonic/percussive source separation. 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**

2.3 TFMBSS

TFMBSS [7] 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 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 [7] 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．

2.4 HPSS に基づく TFMBSS

文献 [8] では，HPSS で算出される時間周波数マスクを TFMBSS に導入した BSS が提案された．本手法のブロック図を Fig. 1 に示す．本手法では，TFMBSS の最適化反復中に，中間変数 $\mathbf{z}$ に対して HPSS を適用する．その結果から新たな時間周波数マスクを生成し，再び TFMBSS で最適化することを繰り返す．

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 $\boldsymbol{Z}_H \in \mathbb{C}^{I \times J}$ 及び $\boldsymbol{Z}_P \in \mathbb{C}^{I \times J}$ と表記すると，HPSS の分離信号の初期値は次式で与える．

$$|\boldsymbol{H}| = |\boldsymbol{Z}_H|, \qquad |\boldsymbol{P}| = |\boldsymbol{Z}_P| \tag{12}$$

ここで，行列に対する演算 $|\cdot|$ は要素毎の絶対値を示す．HPSS の観測信号 $|\boldsymbol{B}|$ は式 (7) で定義される．式 (12) を用いて，式 (8) 及び (9) を反復的に計算することで，中間変数 $\mathbf{z}$ を初期値とした HPSS が可能となる．得られた $\boldsymbol{H}$ と $\boldsymbol{P}$ の推定結果から，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．

$$[\mathcal{M}_H]_{ij} = \left(\frac{|h_{ij}|^2}{|h_{ij}|^2 + |p_{ij}|^2} \right)^{\frac{1}{2}} \tag{13}$$

$$[\mathcal{M}_P]_{ij} = \left(\frac{|p_{ij}|^2}{|h_{ij}|^2 + |p_{ij}|^2} \right)^{\frac{1}{2}} \tag{14}$$

ここで，$\mathcal{M}_H \in \mathbb{R}^{I \times J}_{[0,1]}$ 及び $\mathcal{M}_P \in \mathbb{R}^{I \times J}_{[0,1]}$ はそれぞれ調波音と打撃音の成分を強調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．

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 $\mathcal{M}_{\mathrm{old}}$ とのスムージングを施す．

$$\mathcal{M} = \mathcal{M}^{\beta} \odot {\mathcal{M}_{\mathrm{old}}}^{\beta_{\mathrm{old}}} \tag{15}$$

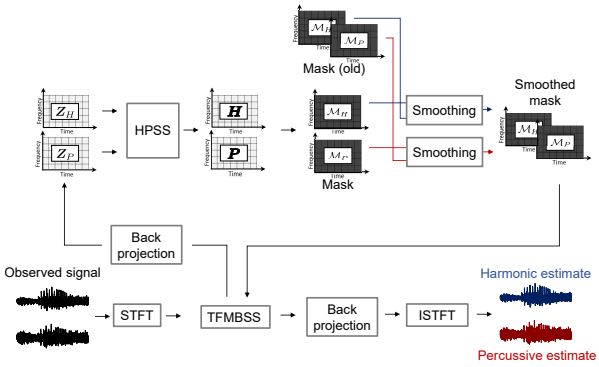

Fig. 1 Block diagram of conventional HPSS-based TFMBSS.

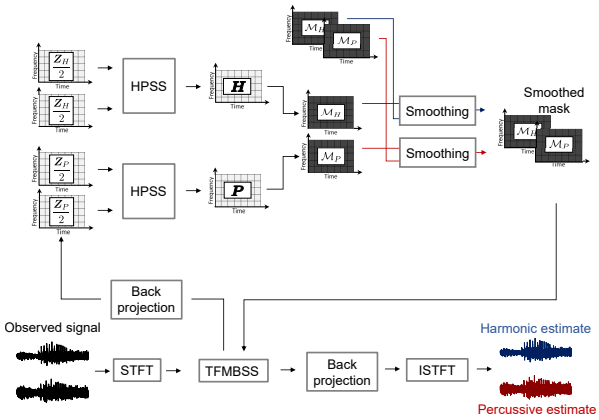

Fig. 2 Block diagram of proposed HPSS-based TFMBSS.

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，$\beta + \beta_{\mathrm{old}} = 1$ を満たす．式 (15) の処理を $\mathcal{M}_H$ 及び $\mathcal{M}_P$ のそれぞれに施す．スムージング後のマスクは 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 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 [10] を適用し，周波数毎のスケールを補正している．

3 提案手法

3.1 動機

従来手法では，中間変数 $\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ をそれぞれ $\boldsymbol{H}$ と $\boldsymbol{P}$ の初期値とし，さらに $\boldsymbol{Z}_H + \boldsymbol{Z}_P$ を観測とする HPSS を用いて，調波音及び打撃音をそれぞれ強調する時間周波数マスクを生成していた．一方で，HPSS による音源強調をフィルタとして捉え，調波音成分 $\boldsymbol{Z}_H$ の中に残留する打撃音成分や，逆に打撃音成分 $\boldsymbol{Z}_P$ 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求めるアプローチを考えることもできる．この場合，従来の時間周波数マスクよりも，残留する干渉音成分を抑圧する能力が高いことが期待できる．本稿では，上記のアルゴリズムを構築し，従来手法と比較することで，提案手法の優位性を確認する．

3.2 HPSS の排他的マスキングに基づく TFMBSS

本稿における提案手法のブロック図を Fig. 2 に示す．提案手法では，2.4 節で述べた従来手法と同様に，TFMBSS の時間周波数マスクを HPSS で生成することを考える．但し，式 (13) 及び (14) によるマスクの生成方法ではなく，より排他的な（他音源を抑圧するような）時間周波数マスクを生成する手法を用いる．

具体的には，Fig. 2 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，2 つの独立な HPSS に観測信号として $|\boldsymbol{B}^{(\boldsymbol{Z}_H)}| = |\boldsymbol{Z}_H|$ 及び $|\boldsymbol{B}^{(\boldsymbol{Z}_P)}| = |\boldsymbol{Z}_P|$ のように与える．また，それぞれの HPSS における初期値は次式で与える．

$$\left|\boldsymbol{H}^{(\boldsymbol{Z}_H)}\right| = \left|\boldsymbol{P}^{(\boldsymbol{Z}_H)}\right| = \left|\frac{\boldsymbol{Z}_H}{2}\right| \tag{16}$$

$$\left|\boldsymbol{H}^{(\boldsymbol{Z}_P)}\right| = \left|\boldsymbol{P}^{(\boldsymbol{Z}_P)}\right| = \left|\frac{\boldsymbol{Z}_P}{2}\right| \tag{17}$$

即ち，$\boldsymbol{Z}_H$ 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 4 種類の信号を推定する．調波音成分と打撃音成分の時間周波数マスクはそれぞれ次式で定義する．

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

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

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．従って，中間変数 $\boldsymbol{Z}_H$ 内に残留する打撃音成分をさらに抑圧し，同様に中間変数 $\boldsymbol{Z}_P$ 内に残留する調波音成分をさらに抑圧するような排他的なマスクを生成できる．

このようにして求まる時間周波数マスクに対し，従来手法と同様に式 (15) によるスムージングを施す．これを新たな時間周波数マスクとして TFMBSS に返す．従来手法と同様に，HPSS を行う前にはプロジェクションバック法を適用し，$\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ 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 [11] 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 [12] 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 [13]（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）[14] の改善量を用いた．その他の実験条件を Table 1 に示す．

4.2 HPSS の反復回数に対する性能の変化

従来手法及び提案手法では，TFMBSS を 1 回反復する毎に，HPSS の更新式（式 (8) 及び (9)）を反復計算している．このときの HPSS の反復回数を変化させることによる SDR 改善量の変化を調査した．

Table 1 Experimental conditions

Window function in STFT	Hann window
Window length in STFT	128 ms
Shift length in STFT	64 ms
Parameters in HPSS	$\gamma_H = 1.02$ $\gamma_P = 1.01$
Parameters in TFMBSS	$\alpha = 0.25$ $\mu_1 = \mu_2 = 1.0$
# of iterations in BSS	500

Table 2 Average SDR improvement for convention and proposed methods with various HPSS iterations

# of iterations in HPSS	Average SDR improvement [dB]	
	Conventional method	Proposed method
1	**7.58**	8.29
3	7.40	10.40
5	7.31	10.87
7	7.25	10.98
9	7.22	11.08
11	7.20	10.79
13	7.20	11.09
15	7.19	**11.29**

従来手法と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，表における両手法の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が良い設定値である（次節の実験結果に基づく）．従来手法では HPSS の反復回数が少ないほど高い性能を示したが，提案手法では逆に HPSS の反復回数が多いほど高い性能を示した．提案手法は，HPSS で求めるべき変数 $\boldsymbol{H}$ 及び $\boldsymbol{P}$ が毎回式 (16) 及び (17) で初期化されるため，従来手法と比較して HPSS での十分な反復計算が必要になる．以降の実験では，この結果に基づき，従来手法の HPSS の反復回数を 1 回，提案手法の HPSS の反復回数を 15 回と設定する．

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．従来手法及び提案手法において，β_{old} 及び β のみを変化させた場合の反復毎の SDR 改善量の例（song no. 18）をそれぞれ Figs. 3 及び 4 に示す．但し，常に $\beta_{\text{old}} + \beta = 1$ である．さらに，従来手法及び提案手法における全 20 曲の平均 SDR 改善量を Table 3 に示す．両手法共に β_{old} を低く設定した場合（スムージングを弱くした場合），SDR 改善量の推移が安定せず収束値も低くなる．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），推移は安定するが収束が遅れている．従来手法と提案手法を比較すると，従来手法では β_{old} の値を小さく設定しても比較的安定した推移が見られるが，提案手法と比べて収束値が低い．提案手法では，β_{old} の値が低いと安定しないが，収束値は従来手法よりも高い．最終的な収束値と安定性のトレードオフを考慮すると，両手法共に $\beta_{\text{old}} = 0.75$ 及び $\beta = 0.25$ が最適である．

4.4 種々の BSS との性能比較

最後に，HPSS 単体（単一チャネル），IVA，及び ILRMA との性能比較を行う．但し，HPSS に基づく TFMBSS は従来手法及び提案手法ともに $\beta_{\text{old}} = 0.75$ 及び $\beta = 0.25$ としている．これらの手法の実験条件を Table 4 に示す．Table 5 は，データセット 20 曲全

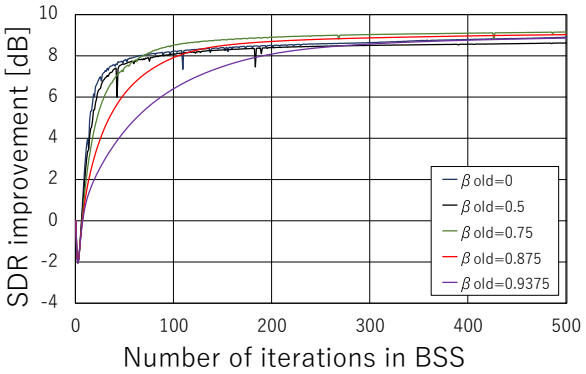

Fig. 3 Example of convergence behaviors of conventional method with various β_{old} and β (song no. 18).

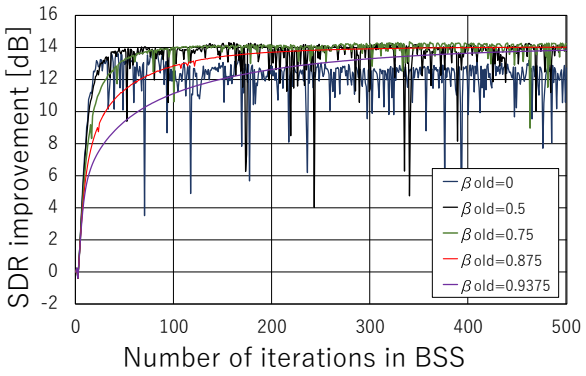

Fig. 4 Example of convergence behaviors of proposed method with various β_{old} and β (song no. 18).

てにおける全手法の平均 SDR 改善量の比較である．これを見ると，提案手法は他のどの手法よりも高い性能を示しており，調波音と打撃音の多チャネル音源分離に特化した高性能な BSS であることが確認できる．

5 まとめ

本稿では，HPSS の音源モデルに基づく TFMBSS として，より排他的な時間周波数マスクを求めるアルゴリズムを新たに提案した．HPSS の反復回数及びマスクのスムージング度合いに対する性能の変化を実験的に確認し考察した．提案手法は，従来の BSS 手法よりも大幅な性能改善が得られることを示した．

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

Table 3 Average SDR improvement for conventional and proposed methods with various β_{old} and β

Parameter β_{old}	Average SDR improvement [dB]	
	Conventional method	Proposed method
0	**8.14**	7.40
0.5	7.55	10.14
0.75	7.58	**11.29**
0.875	7.36	11.27
0.9375	6.82	11.01

Table 4 Experimental conditions for other BSS methods

Window function in STFT	Hann window
Window length in STFT	256 ms
Shift length in STFT	128 ms
# of iterations in simple HPSS	15
# of iterations in IVA	30
# of iterations in ILRMA	100
# of bases in ILRMA	30

Table 5 Average SDR improvement for each method

Method	Average SDR improvement [dB]
Simple HPSS	5.92
IVA	7.92
ILRMA	9.20
Conventional method	7.58
Proposed method	**11.3**