時変な全極モデルを用いた基底変形型教師あり NMF による音楽信号分離*

◎中嶋広明（東大）, 北村大地（総研大）, 高宗典玄（東大）, 小山翔一（東大）, 猿渡洋（東大）,
小野順貴（NII/総研大）, 高橋祐（ヤマハ）, 近藤多伸（ヤマハ）

1 はじめに

近年、低ランク近似手法である非負値行列因子分解 (nonnegative matrix factorization: NMF) による音源分離技術が活発に研究されている。NMF とは、スペクトログラムを非負の基底行列とアクティベーション行列の積に分解する手法である。特に、モノラル・フォーマットで収録・再現される音楽信号の分離において、NMF は有力な手法の一つになっている [1]。

NMF による音源分離手法は、大きく分けて教師なしと教師ありに分けられる。教師なし NMF では、訓練データを与えないまま NMF により音源分離を達成する [2]。教師あり NMF (supervised NMF: SNMF) では、目的音のサンプルを訓練データとして利用することで音源分離を達成する [3]。しかし SNMF において、事前に学習する教師音と実際に分離したい目的音の音色が異なる場合に分離の性能が著しく劣化するという問題点がある。

この問題を解決するために、我々は混合音中に含まれる目的音へ適応して教師基底を時不変に変形する NMF を用いた分離アルゴリズムを提案した [4]。本稿では時変な全極モデルに基づく教師基底の変形と、妨害音成分を表現することを防止する識別的な方法を組み合わせた手法を提案し分離性能が改善することを確認した。

2 従来手法

2.1 SNMF 音源分離の概要

SNMF は事前学習と観測混合音分離の 2 段階から構成される。以降で詳細を述べる。

まず、事前学習によって教師基底を得る。式 (1) のように、抽出したい目的音による単独の演奏音 $\boldsymbol{Y}_{\text{target}}$ を教師音として NMF により基底行列とアクティベーション行列に分解することで、当該目的音に関する教師基底を得る。

$$\boldsymbol{Y}_{\text{target}} \simeq \boldsymbol{F}\boldsymbol{G} \tag{1}$$

$\boldsymbol{Y}_{\text{target}}$ は Ω 行 T_1 列の非負値行列で、ここでは振幅スペクトログラムを表すものとする。$\boldsymbol{F}$ は Ω 行 K 列の非負値行列で一般的に基底行列と呼ばれる。$\boldsymbol{G}$ は K 行 T 列の非負値行列で、$\boldsymbol{F}$ の各基底ベクトルの励起情報を持っていることからアクティベーション行列と呼ばれる。ここで、K は行列 $\boldsymbol{F}$ の基底数である。

SNMF の観測混合音の分離は、事前学習で得た基底行列 $\boldsymbol{F}$ を用いて次式のように書くことができる。

$$\boldsymbol{Y}_{\text{mix}} \simeq \boldsymbol{F}\boldsymbol{G} + \boldsymbol{H}\boldsymbol{U} \tag{2}$$

$\boldsymbol{Y}_{\text{mix}}$ は Ω 行 T_2 列の非負値行列で、ここでは振幅スペクトログラムである。この分離は事前学習で得た基底行列 $\boldsymbol{F}$ を与えた上で求めているため、理想的には $\boldsymbol{F}\boldsymbol{G}$ が観測混合音中の抽出したい目的音成分を表し、$\boldsymbol{H}\boldsymbol{U}$ が妨害音を表す。

式 (2) の分解は、以下の最小化問題を解くことに帰着する。

$$\min_{\boldsymbol{G},\,\boldsymbol{H},\,\boldsymbol{U}} \mathcal{D}_{\text{KL}}(\boldsymbol{Y}|\boldsymbol{F}\boldsymbol{G} + \boldsymbol{H}\boldsymbol{U}) \tag{3}$$

ただし $\mathcal{D}_{\text{KL}}$ は一般化 KL ダイバージェンスであり、次のように定義する。

$$\mathcal{D}_{\text{KL}}(y|x) = y\log\frac{y}{x} + x - y \tag{4}$$

式 (3) の最小化問題は、補助関数法により求めることができる。

2.2 加法型基底変形 NMF

2.1 節で述べた SNMF では、事前学習で用いる教師音と観測混合音中の目的音の間で音色が異なる場合に、大きく分離性能が低下するという欠点がある。そこで、目的音に対し事前学習で得た基底行列を変形させることで分離性能の向上を目指す手法として、加法型基底変形 NMF [5] が提案された。

加法型基底変形 NMF における観測混合音の分離は、事前学習で得た基底行列 $\boldsymbol{F}$ を用いて次式のように書くことができる。

$$\boldsymbol{Y}_{\text{mix}} \simeq (\boldsymbol{F} + \boldsymbol{D})\boldsymbol{G} + \boldsymbol{H}\boldsymbol{U} \tag{5}$$

$\boldsymbol{D}$ は F とアクティベーション行列 $\boldsymbol{G}$ を共有する Ω 行 K 列の基底行列である。この分解では、$\boldsymbol{F}$ では表現しきれない目的音の基底を $\boldsymbol{D}$ で表現している。$\boldsymbol{D}$ は非負値行列ではないが、$\boldsymbol{F}+\boldsymbol{D}$ が非負値になるよう適当な制約を加えるものとする。

しかしこの手法には、3 つ以上の重みパラメータがあるため調整が困難なこと、各変数の初期値依存性が極めて強いこと、線形時不変な変形ではないなどの問題点が存在する。

2.3 統計的音声スペクトル推定器を用いたポストフィルタ

一般化 minimum mean-square error short-time spectral amplitude (MMSE-STSA) 推定器 [6] は、目的音を表す事前分布の基で、真の目的音振幅スペクトルとその推定値との平均二乗誤差を最小化するスペクトルゲインを求める。このスペクトルゲインを観測信号スペクトルに乗算することで推定目的音が得られ、音源分離が達成できる。一般化 MMSE-STSA 推定器における推定目的音を $\hat{s}(f,\tau)$、観測混合音を $x(f,\tau)$ とすると、スペクトルゲイン $G(f,\tau)$ は次のように与えられる。

---

* "Music Signal Separation Using Supervised NMF with Time-Variant All-Pole-Model-Based Basis Deformation," by Hiroaki Nakajima (The University of Tokyo), Daichi Kitamura (SOKENDAI), Norihiro Takamune (The University of Tokyo), Shoichi Koyama (The University of Tokyo), Hiroshi Saruwatari (The University of Tokyo), Nobutaka Ono (NII/SOKENDAI), Yu Takahashi (Yamaha Corp.), Kazunobu Kondo (Yamaha Corp.).

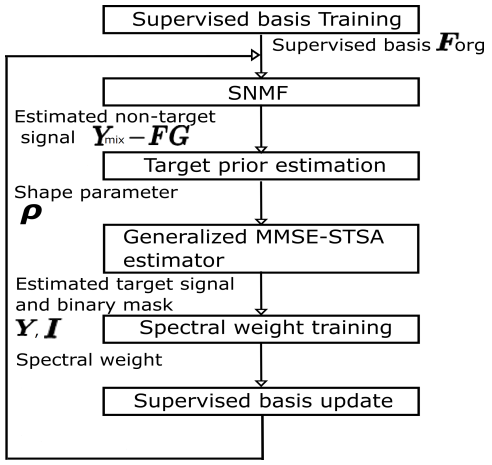

Fig. 1　一般化 MMSE-STSA 推定器を用いた時不変全極モデルによる教師基底変形手法のブロック図

$$\tilde{s}(f,\tau) = G(f,\tau)x(f,\tau) \tag{6}$$

$$G(f,\tau) = \frac{\sqrt{v(f,\tau)}}{\tilde{\gamma}(f,\tau)} \cdot \left(\frac{\Gamma(\rho+0.5)}{\Gamma(\rho)} \cdot \frac{\Phi(0.5-\rho,1,-v(f,\tau))}{\Phi(1-\rho,1,-v(f,\tau))}\right)^{1/\beta} \tag{7}$$

ここで $\Gamma(\cdot)$ はガンマ関数、$\Phi(a,b;k) = F_1(a,b;k)$ は合流型超幾何関数、β は振幅圧縮パラメータ、ρ は事前分布であるカイ分布における形状母数を表す。また、$v(f,\tau)$ は事前 SN 比 $\tilde{\epsilon}(f,\tau)$、事後 SN 比 $\tilde{\gamma}(f,\tau)$ を用いて次のように与えられる。

$$v(f,\tau) = \tilde{\gamma}(f,\tau)\tilde{\epsilon}(f,\tau)\Big(1+\tilde{\epsilon}(f,\tau)\Big)^{-1} \tag{8}$$

一般化 MMSE-STSA 推定器では、$\tilde{\gamma}(f,\tau)$ の計算に妨害音のパワースペクトルが必要になるが、本研究では基底変形前の教師基底による SNMF によって得られる妨害音 $\boldsymbol{Y}_{\text{mix}} - \boldsymbol{FG}$ を用いる。

2.4　一般化 MMSE-STSA 推定器の出力を用いた時不変な全極モデルによる教師基底変形

従来の基底変形型 SNMF では、教師基底の変形と信号分離を同時に行うため最適化が困難であった。そこで我々は、SNMF において一般化 MMSE-STSA 推定器の出力値を用いることで信号分離とは別に教師基底変形を行い最適化を容易にする手法を提案した [4]。

処理の全体像を Fig. 1 に示す。第一に現在得ている教師基底を用いて SNMF を行う。得られた妨害音を初期値として一般化 MMSE-STSA 推定器を適用し、時間周波数領域で推定目的音及び各成分の確信度をしめすバイナリマスクが得られる。最後に以下で提案する全極モデルによる教師基底変換を施し、変形された教師基底を新しい教師基底とする。これらの処理を任意の反復回数だけ行い、得られる新しい教師基底による SNMF の推定目的音を最終的な出力とする。教師基底の変形成分は次のように学習する。

$$\boldsymbol{I} \circ \boldsymbol{Y} \simeq \boldsymbol{I} \circ (\boldsymbol{A}\boldsymbol{F}_{\text{org}}\boldsymbol{G}) \tag{9}$$

ここで $\boldsymbol{Y}$ は前の反復で得られる一般化 MMSE-STSA 推定器の出力、$\boldsymbol{I}$ は妨害音との重複の少ない成分（すなわち目的音成分であることが確かな成分）を $\boldsymbol{Y}$ から抜き出すバイナリマスク、$\boldsymbol{F}_{\text{org}}$ は教師音を用いて学習した教師基底、$\boldsymbol{A}$ は対角成分に全極モデルによるスペクトル重みを持つ行列、$\circ$ はアダマール積を表す。$\boldsymbol{I}$ は一般化 MMSE-STSA 推定器で得られるスペクトルゲイン $G(f,\tau)$ を閾値処理して得る [7]。

3　提案手法

3.1　時不変な全極モデルによる教師基底変形法の課題と提案手法の概要

2.4 節で提案したモデルは、目的音の分離に適するように教師基底を線形時不変に変形するものであった。つまり時間変化せず、すべての基底に共通の変形係数を施すモデルである。しかし楽器が音を発生させる物理現象は時不変ではなく、鳴り始めから鳴り終わりまでの間で時間的に変化する [8]。したがって、異なる物理状態を表す基底には異なる変形を施したほうが、より目的音の分離に適した基底へ変形することが出来ると考えれらる。つまり分離性能の向上には教師基底の時変な変形が必要である。

一方で、教師基底の変形の自由度を向上させることは、妨害音を表しやすくなる基底へ変形させる可能性を高めることと等価である。したがって変形の自由度をある程度許しつつ、自由度を合理的に制限する制約が求められる。

そこで本章では、一般化 MMSE-STSA 推定器の出力値を用いて楽器音を鳴り始めと鳴り終わりの 2 種類の状態で別々の変形を施す時変な教師基底変形手法と、妨害音からの分離を考慮した正則化項を加える識別的な教師基底変形手法を提案する。

3.2　時変な全極モデルによる教師基底の変形

本節では、一般化 MMSE-STSA 推定器の出力値を用いて楽器音を鳴り始めと鳴り終わりの 2 種類の状態で別々の変形を施す変形手法を提案する。以下では、鳴り始めと鳴り終わりを表現する基底を得る方法について述べる。最初に教師音を、鳴り始め部分のみを切り出した音と鳴り終わり部分のみを切り出した音へ分割する。次にこれらを STFT することによってスペクトログラムを求め、教師音全体から得られる教師基底 $\boldsymbol{F}_{\text{org}}$ を固定してアクティベーションを得る。このアクティベーションを行方向に総和を取ると、各楽器音の物理状態における各基底の励起情報が得られる。最後に鳴り始めにおける各基底の励起度合いと鳴り終わりにおける各基底の励起度合いを入力とした k 平均法によって元の教師音から得た教師基底 $\boldsymbol{F}_{\text{org}}$ を 2 つに分割することで、主に鳴り始めを表現する基底 $\boldsymbol{F}_1$ と主に鳴り終わりを表現する基底 $\boldsymbol{F}_2$ が得られる。

次に、このクラスタリングされた教師基底を用いて、以下の教師基底の変形成分の学習を行う。

$$\boldsymbol{I} \circ \boldsymbol{Y} \simeq \boldsymbol{I} \circ (\boldsymbol{A}\boldsymbol{F}_1\boldsymbol{G}_1 + \boldsymbol{B}\boldsymbol{F}_2\boldsymbol{G}_2) \tag{10}$$

ここで $\boldsymbol{F}_1$ は教師基底行列 $\boldsymbol{F}_{\text{org}}$ において鳴り始めの部分を主に表現する基底行列、$\boldsymbol{F}_2$ は教師基底行列 $\boldsymbol{F}_{\text{org}}$ において鳴り終わりの部分を主に表現する基底行列、$\boldsymbol{A}$ は鳴り始めを表す教師基底を変形するための対角成分に全極モデルによるスペクトル重みを持つ行列、$\boldsymbol{B}$ は鳴り終わりを表す教師基底を変形するための対角成分に全極モデルによるスペクトル重みを持つ行列を表

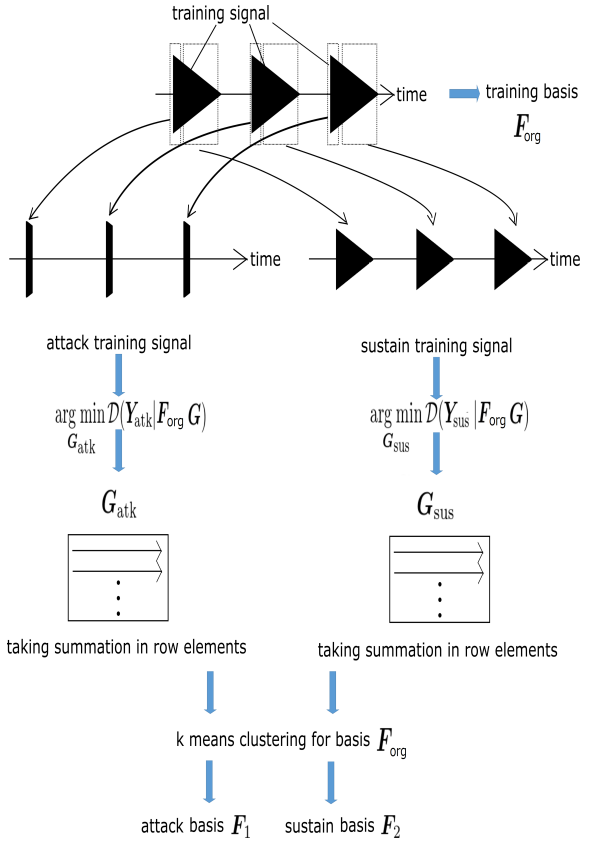

Fig. 2　鳴り始めのみからなる教師音と鳴り終わりのみからなる教師音の作成方法の概略図

す。$\boldsymbol{A}, \boldsymbol{B}$ の対角成分は全極モデルで与える。

一般化 KL ダイバージェンス基準の時変な教師基底の変形の目的関数 $\mathcal{J}$ は次のように書ける。

$$\mathcal{J} = \sum_{\omega,t} i_{\omega,t}\Big\{-y_{\omega,t} + \frac{\sum_k f_{\omega,k,1} g_{k,t,1}}{|A_\omega|} + \frac{\sum_l f_{\omega,l,2} g_{l,t,2}}{|B_\omega|} + y_{\omega,t} \log \frac{y_{\omega,t}}{\sum_k f_{\omega,k,1} g_{k,t,1}/|A_\omega| + \sum_l f_{\omega,l,2} g_{l,t,2}/|B_\omega|}\Big\} \tag{11}$$

この目的関数の最適化は補助関数法を用いて解くことが出来る。

3.3　識別的な教師基底の変形

提案手法では教師基底の変形に一般化 MMSE-STSA 推定器の出力を利用している。したがって推定目的音 $\boldsymbol{Y}$ には妨害音成分も含まれているので、妨害音成分を表しやすい基底へ変形する可能性がある。そこで全教師あり（目的音・妨害音教師音を利用可能）NMF における妨害音を考慮した教師基底学習 [9] を拡張し、半教師あり NMF における教師基底変形問題で適用できる手法を提案する。このような学習を識別的学習という。まず、この問題を 2 段階最適化として定式化すると以下のようになる。

$$\boldsymbol{A}, \boldsymbol{B} = \arg\min_{\boldsymbol{A},\boldsymbol{B}} \mathcal{D}\Big(\boldsymbol{I} \circ \boldsymbol{Y} | \boldsymbol{I} \circ (\boldsymbol{A}\boldsymbol{F}_1\boldsymbol{G}_{1s} + \boldsymbol{B}\boldsymbol{F}_2\boldsymbol{G}_{2s})\Big) \tag{12}$$

$$\text{s.t. } \boldsymbol{G}_{1s}, \boldsymbol{G}_{2s} = \arg\min_{\boldsymbol{G}_1,\boldsymbol{G}_2,\boldsymbol{H},\boldsymbol{U}} \mathcal{D}\Big(\boldsymbol{I} \circ \boldsymbol{Y}_{mix} | \boldsymbol{I} \circ (\boldsymbol{A}\boldsymbol{F}_1\boldsymbol{G}_1 + \boldsymbol{B}\boldsymbol{F}_2\boldsymbol{G}_2 + \boldsymbol{H}\boldsymbol{U})\Big) \tag{13}$$

この最適化問題は、$\boldsymbol{Y}_{mix}$ を $\boldsymbol{A}\boldsymbol{F}_1\boldsymbol{G}_1 + \boldsymbol{B}\boldsymbol{F}_2\boldsymbol{G}_2 + \boldsymbol{H}\boldsymbol{U}$ として良く表現できるアクティベーション $\boldsymbol{G}_{1s}, \boldsymbol{G}_{2s}$ を基底の変形項 $\boldsymbol{A}, \boldsymbol{B}$ の関数として求めた中で、推定目的音 $\boldsymbol{Y}$ を良く表現する基底の変形項を求めるという意味を持つ。この 2 段階最適化を解くことは困難なので、近似解を与える反復アルゴリズムを以下のように提案する。

Step 1 : Initialization

$$\boldsymbol{A}, \boldsymbol{B} = \arg\min_{\boldsymbol{A},\boldsymbol{G}_1,\boldsymbol{B},\boldsymbol{G}_2} \mathcal{D}\Big(\boldsymbol{I} \circ \boldsymbol{Y} | \boldsymbol{I} \circ (\boldsymbol{A}\boldsymbol{F}_1\boldsymbol{G}_1 + \boldsymbol{B}\boldsymbol{F}_2\boldsymbol{G}_2)\Big) \tag{14}$$

Step 2 : Modeling of Mixture $\boldsymbol{Y}_{mix}$

$$\boldsymbol{G}_1, \boldsymbol{G}_2 = \arg\min_{\boldsymbol{G}_1,\boldsymbol{G}_2,\boldsymbol{H},\boldsymbol{U}} \mathcal{D}\Big(\boldsymbol{I} \circ \boldsymbol{Y}_{mix} | \boldsymbol{I} \circ (\boldsymbol{A}\boldsymbol{F}_1\boldsymbol{G}_1 + \boldsymbol{B}\boldsymbol{F}_2\boldsymbol{G}_2 + \boldsymbol{H}\boldsymbol{U})\Big) \tag{15}$$

Step 3 : Modeling of Target $\boldsymbol{Y}$

$$\boldsymbol{A}, \boldsymbol{B} = \arg\min_{\boldsymbol{A},\boldsymbol{B}} \mathcal{D}\Big(\boldsymbol{I} \circ \boldsymbol{Y} | \boldsymbol{I} \circ (\boldsymbol{A}\boldsymbol{F}_1\boldsymbol{G}_1 + \boldsymbol{B}\boldsymbol{F}_2\boldsymbol{G}_2)\Big) \tag{16}$$

Return to Step 2

このアルゴリズムは、推定目的音 $\boldsymbol{Y}$ を良く表現する基底 $\boldsymbol{A}\boldsymbol{F}_1 + \boldsymbol{B}\boldsymbol{F}_2$ の周辺で $\boldsymbol{Y}_{mix}$ を $\boldsymbol{A}\boldsymbol{F}_1\boldsymbol{G}_1 + \boldsymbol{B}\boldsymbol{F}_2\boldsymbol{G}_2 + \boldsymbol{H}\boldsymbol{U}$ として良く表現できる基底を見つける。

4　評価実験

4.1　実験条件

提案手法の有効性を確認するために、音楽信号を対象とした分離評価実験を行った。まず目的音として Microsoft GS Wavetable SW Synth で作成した Oboe (Ob.), Piano (Pf.), Trombone (Tb.) の 3 種類を用意した（それぞれの楽譜は [6] を参照されたい）。それらの中から 2 種類の楽器音を選び、等パワーで混合させることにより観測音（分離対象音）を作成した。また教師音として Garritan Personal Orchestra で作成した Ob., Pf., Tb. を使用する。この教師音は、観測音中の各楽器音の音域を含む範囲で、2 オクターブにわたり半音階ずつ上昇する 24 音で構成されている。また教師音および目的音はいずれもサンプリング周波数 44.1 kHz であり、窓長 4096 点の矩形窓を 512 点でシフトさせてスペクトログラムを作成した。また事前学習における学習基底 $\boldsymbol{F}$ の数は 100、信号分離における妨害音基底 $\boldsymbol{H}$ の数は 30 とした。一般化 MMSE-STSA 推定における振幅圧縮パラメータ β は 1.0 とした。また識別的な教師基底変形において、Step 2 と Step 3 の反復更新回数を 1 回とし、Step 2 と Step 3 の交互更新は 10

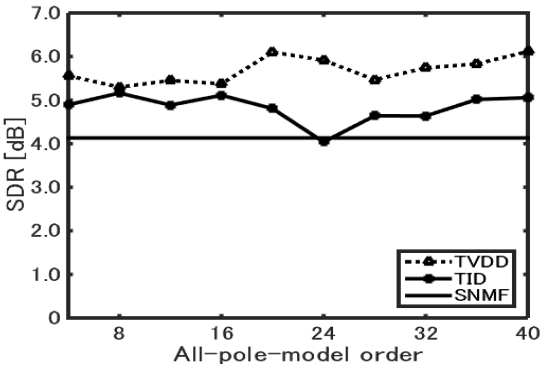

Fig. 3　Pf. と Ob. の混合音から Pf. を分離抽出する場合における SDR

回とする。これは目的音を良く表す基底から過度に離れてしまうことを防止するためである。

客観評価指標として BSS EVAL TOOLBOX [10] による source-to-distortion ratio (SDR) を用いる。SDR は目的音と妨害音の分離度合いと、一連の信号処理によって生じた目的音の線形歪み及び非線形歪みを考慮した推定目的音の品質を表すものであり、値が大きくなるほど精度がよいことを示す。

実験の概要は以下のとおりである。ここでは、混合された 2 つの楽器音から 1 つの目的音を分離する問題を考える。分離したい音の教師音は与えられているが、音色が異なっている（なぜならば異なる MIDI 音源を使ってそれぞれ生成されているから）。提案手法においては、基底変形の際に用いる全極モデルの次数を共に 1 から 40 まで 4 次刻みとし、3.2 節と 3.3 節で示した一連の反復処理を 8 回繰り返した後に得られる変形した教師基底を用いた SNMF による SDR を計測した。比較対照手法は、2.1 節で説明した（基底変形を含まない）SNMF (SNMF)、2.4 節で説明した時不変な全極モデルによる教師基底変形 (time-invariant deformation: TID)、及び提案手法 (time-variant and discriminative deformation: TVDD) とする。

4.2　実験結果

Figure 3 は、Pf. と Ob. の楽器音の組み合わせから Pf. の音を分離するという問題に関して、SNMF、TID、TVDD を比較したものである。提案手法が SNMF と TID を上回っていることがわかる。

Table 1 は従来の SNMF と、提案手法において全極モデルの次数を 1 から 40 まで変化させた時の SDR の値が最大になる次数とその SDR の値を各楽器音の組み合わせに対して記述したものである。これは、組み合わせパタン（例えば Ob. & Pf.）のうち、前方に表記されている楽器音（前記例で言えば Ob.）を分離する実験の結果を表している。TVDD において、SNMF および TID の SDR を上回る次数が存在することがわかる。ただし、共通の全極モデル次数において、常に TVDD が TID を上回っているわけではなく、個々の手法において最大性能を示す次数は異なっていた。最適次数の決定問題は、今後の課題だと言える。

Table 1　各楽器混合パタンにおける分離抽出音 SDR の最大値 [dB]

	SNMF	TID	TVDD
Ob. & Pf.	6.655	6.740	**7.004**
Ob. & Tb.	2.374	2.811	**2.934**
Pf. & Ob.	4.132	5.162	**6.115**
Pf. & Tb.	3.077	4.482	**4.542**
Tb. & Ob.	0.661	2.387	**2.831**
Tb. & Pf.	2.942	3.852	**4.415**

5　おわりに

本稿では、一般化 MMSE-STSA 推定器の出力を用いた時変な全極モデルによる教師基底変形と妨害音成分を表現することを防ぐ識別的な手法を組み合わせることにより、SNMF や時不変な全極モデルによる教師基底変形手法よりも優れた分離を達成できることを確認した。

References

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

[2] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino and S. Sagayama, “Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints,” *Proc. ICASSP*, pp.5365–5368, 2012.

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

[4] 中嶋広明, 北村大地, 高宗典玄, 小山翔一, 猿渡洋, 小野順貴, 高橋祐, 近藤多伸, ”全極モデルを用いた基底変形型教師あり NMF による音楽信号分離,” 日本音響学会 *2015* 年秋季研究発表会, no.3-6-7, pp.573-576, 福島, 2015 年 9 月.

[5] D. Kitamura, H. Saruwatari, K. Shikano, K. Kondo, Y. Takahashi, “Music signal separation by supervised nonnegative matrix factorization with basis deformation,” *Proc. IEEE 18th International Conference on Digital Signal Processing (DSP2013)*, no.T3P(C)-1, 2013.

[6] Y. Murota, D. Kitamura, S. Nakai, H. Saruwatari, S. Nakamura, K. Shikano, Y. Takahashi, K, Kondo,“Music signal separation based on bayesian spectral amplitude estimator with automatic target prior adaptation,” *Proc. ICASSP2014*, pp.7490–7494, 2014.

[7] 室田勇騎, 北村大地, 小山翔一, 猿渡洋, 中村哲, ”時系列事前分布モデルとスペクトル基底の同時適応を用いたバイノーラル音源分離の実験的評価,” 日本音響学会 *2015* 年春季研究発表会, no.1-10-12, pp.549-552, 東京, 2015 年 3 月.

[8] N. H. Fletcher, T. D. Rossing, 岸憲史, 久保田秀美, 吉川茂, “楽器の物理学,”シュプリンガー・フェアラーク東京, 2002.

[9] F. Weninger, J. Le Roux, J. R. Hershey, S. Watanabe, “Discriminative NMF and its application to single-channel source separation,” *Proc. ISCA Interspeech 2014*, Sep. 2014.

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