

非負値周期成分分析手法による音楽音響信号の音源分離*

©林亜紀 (NTT EV 研), 亀岡宏和 (NTT CS 研), △松林達史, 澤田宏 (NTT EV 研)

1 はじめに

音楽音響信号の多くは、複数の楽器などの音源から構成され、混合信号の分離は音楽構造理解や自動採譜、楽器やピッチの部分的な編集、雑音除去、欠損値補間など多くの場面で必要になる。近年、音響信号処理分野において、自動採譜や音源分離などへの有効性から非負値行列分解 (Non-negative Matrix Factorization, NMF) [1] への注目が高まっている [2]。NMF では、混合信号 (観測信号) のスペクトログラムを非負値行列 \mathbf{Y} で表現した上で、2つの非負値行列 \mathbf{H} , \mathbf{U} の積の形に分解する。これは、各時間窓で観測されたスペクトルを、時間変化する係数行列 \mathbf{U} (アクティベーション) の値でスケールされた基底スペクトル \mathbf{H} の線形和で近似することに該当し、観測行列はランク 1 のスペクトログラム $H_{\omega,k}U_{k,t}$ の和の形で表現される。NMF の重要な特徴として、非負値制約により分解行列がスパースな値を持つことが挙げられる。例えば、 \mathbf{U} には多くの 0 要素が含まれる。この特徴により、各時間窓の観測スペクトルは、多くの場合少数の基底スペクトルの和で表現される。従って、各基底スペクトルには観測信号全体において複数回出現する典型的なスペクトル (多くは頻出する楽音) が抽出され、効率の良い分解表現が可能になる。

しかしながら、音楽スペクトログラムに NMF を適用する際には、2つの課題がある。第 1 に、精度と分離信号の有意性との間にトレードオフが発生する。NMF では $\mathbf{X} = \mathbf{H}\mathbf{U}$ と観測行列 \mathbf{Y} の間の近似誤差を目的関数として定義し、その目的関数を最小化する。基底数を増やすことにより近似誤差は減少するが、基底数を増やしすぎると分解行列の解釈性が下がる。例えば、極端な例として $\mathbf{Y} = \mathbf{H}\mathbf{I}$ や $\mathbf{Y} = \mathbf{I}\mathbf{U}$ などの無意味な分解結果が出力されてしまう場合がある。ここで \mathbf{I} は単位行列を示す。したがって、多くの基底を用いて精度 (少ない近似誤差) を保ちながら、優位性の高い分離を行うためには、非負値性以外の有用な制約を $\mathbf{H}\mathbf{U}$ に課したモデルが必要になる。

NMF の第 2 の課題として、局所的最適解問題が挙げられる。既存の NMF の最適化アルゴリズムの多くは収束が保証されているものの、ある限られた領域においてのみ最適な局所的最適解へ収束してしまう場合が多く、必ずしも全ての領域において最適な大域的最適解が求まるとは限らない。音楽音響信号への適用において、この局所的最適解は、例えば分離信号が聴覚で捉えられる音楽事象に対応しないなど、音楽的に解釈不可能な解となり得る。

上記の問題を解決するには、データの特徴を反映した制約をモデル $\mathbf{H}\mathbf{U}$ や目的関数に組み込むことが考えられる。例えば、音色の類似性を考慮して基底をクラスタリングする枠組みを導入した手法 [3] が提案されている。音楽の時間変化特徴を考慮した制約付 NMF も提案されており、発音が一定時間持続することを仮定したモデル [4] や、「アタック」「ディケイ」

「サステイン」「リリース」などの楽音の状態変化を基底スペクトルに仮定したモデル [5]、ビート (拍) に応じたリズムをアクティベーションに仮定した制約付 NMF [6] が提案されている。

本稿では、音楽音響信号に含まれる、各音符やドラム音の発音タイミングにおける、リズムの周期性を仮定した制約付 NMF を提案する。特にベースやドラムなどの楽器で周期性が想定される。単純かつ単一の周期成分を抽出する手法 [7] が提案されているが、周期性は単一であるとは限らず、複数の音源で異なる複数の周期性が存在する可能性がある。そこで、周期性を想定した一部の基底で、どの音源がいつ発音するかを示すアクティベーションに基底毎に異なる周期性を仮定し、目的関数に制約として加えることにより、周期的な発音パターン抽出を促進する。周期性を仮定する制約を加えた目的関数において、最適解を求めるアルゴリズムを定式化した上で、教師あり・教師なし両方の枠組みにおいて音源分離の性能を評価する。

2 提案モデル

本稿では、提案する周期を考慮したパターン抽出方法を非負値周期成分分析手法 (Non-negative Periodic Component Analysis, NPCA) と呼ぶ。NPCA の目的関数は以下のように定義される。

$$I(\mathbf{H}, \mathbf{U}) + w_1 J(\mathbf{U}, \alpha, \beta) + w_2 L(\alpha, \beta) \quad (1)$$

$w_1, w_2 > 0$ は正規化パラメータである。 $I(\mathbf{H}, \mathbf{U})$ は通常の NMF の目的関数で考慮されるものと同じ、観測行列 \mathbf{Y} と、NMF モデルによる分解行列の積 $\mathbf{X} = \mathbf{H}\mathbf{U}$ との近似誤差である。以下のような NMF モデルを考える。

$$Y_{\omega,t} \simeq X_{\omega,t} = \sum_{k=1}^K H_{\omega,k} U_{k,t} \quad (2)$$

ここで、 k は基底インデックス、 K は基底数を示し、 ω と t はそれぞれ周波数と時間窓を示す。 \mathbf{Y} と \mathbf{X} との間の距離尺度として I ダイバージェンスを用いると、 $I(\mathbf{H}, \mathbf{U})$ は以下のように書ける。

$$I(\mathbf{H}, \mathbf{U}) = \sum_{\omega,t} \left(Y_{\omega,t} \log \frac{Y_{\omega,t}}{X_{\omega,t}} - Y_{\omega,t} + X_{\omega,t} \right) \quad (3)$$

$J(\mathbf{U}, \alpha, \beta)$ はアクティベーション $U_{k,1}, \dots, U_{k,T}$ に周期性を仮定するための制約を表す項である。 T は時間-周波数表現における時間窓数を示す。提案法では、 \mathbf{U} の非負性を保つために、 $U_{k,t}$ の対数を取ったものが、倍音構造を考慮した正弦曲線の重み付きとで近似されること、すなわち

$$\log U_{k,t} \simeq \sum_{m,n} (\alpha_{m,n}^k \cos nP_m t + \beta_{m,n}^k \sin nP_m t) \quad (4)$$

* Non-negative Periodic Component Analysis for Music Source Separation. by HAYASHI, Aki (NTT Service Evolution Laboratories), KAMEOKA, Hirokazu (NTT Communication Science Laboratories), MATSUBAYASHI Tatsushi, SAWADA Hiroshi (NTT Service Evolution Laboratories)

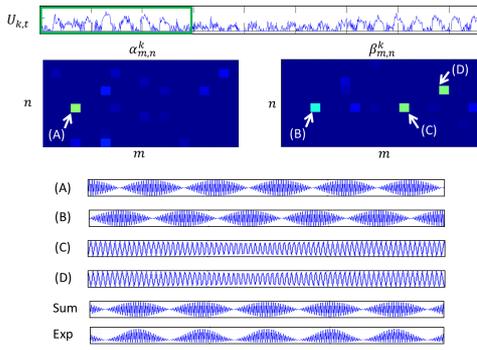


Fig. 1 An example of the estimates of $U_{k,t}$, $\alpha_{m,n}^k$ and $\beta_{m,n}^k$.

と表現できることを仮定する. $\alpha_{m,n}^k, \beta_{m,n}^k \in \mathbb{R}$ が、周期 $1/P_m$ の繰り返しパターンを決定する. ここで、各正弦曲線の周期は基本周波数 P_m の整数倍に限定される形を取っており、 n は各周波数の n 倍音を表現するための変数である. この制約により、 $\sum_n (\alpha_{m,n}^k \cos nP_m t + \beta_{m,n}^k \sin nP_m t)$ がちょうど周波数 P_m の周期構造を持つことが保証される. 以上により、 $\log U_{k,t}$ は M 個の周期関数の和で表現される. なお、式 (4) は \mathbf{U} を正弦曲線の重み付き和の指数の形で近似していると捉えられる.

周期 $\{P_m\}$ について、例えば $P_m = \frac{T}{\log M} \log m + 1.0$, などと定義することが考えられる. また、倍音を表す定数の集合 $\{n\}$ について、 $\{1, 2, \dots, 9, 10\}$ などの定義が考えられる. 上述の例では、想定できる周期の数 M は 20 となる.

提案モデルでは、自動的に $U_{k,t}$ を最適に表現する周期を推定することができる. 図 1 に、ある短い音楽データにおける $U_{k,t}$ と、 $U_{k,t}$ の太枠で囲まれた時間窓内に対応する $\alpha_{m,n}^k, \beta_{m,n}^k$ の推定例を示す. 例においては、(A)-(D) で示されるいくつかの m と n の組み合わせに該当する $\alpha_{m,n}^k$ と $\beta_{m,n}^k$ の値が他の m や n に比べて、大きい値を持つと推定された. 下の周期関数の図は、抽出された m, n に対応する $\cos nP_m t$ または $\sin nP_m t$ の \mathbf{U} の太枠内の時間窓における離散曲線を示す. Sum と書かれた曲線は、(A)-(D) に該当する曲線の重み付き和を示し、Exp と書かれた尖った曲線は、Sum の指数を取ったものである. Exp の曲線は実際に推定された $U_{k,t}$ に近い形をしていることが分かる.

提案手では、アクティベーション \mathbf{U} に周期性を仮定した目的関数 $J(\mathbf{U}, \alpha, \beta)$ を、式 (4) の両辺間のユークリッド距離を用いて、以下のように定義する.

$$J(\mathbf{U}, \alpha, \beta) = \sum_{k,t} |\log U_{k,t} - f_{k,t}|^2 \quad (5)$$

ここで、以下のように置いた.

$$f_{k,t} = \sum_{m,n} \left(\alpha_{m,n}^k \cos nP_m t + \beta_{m,n}^k \sin nP_m t \right) \quad (6)$$

式 (4) で \mathbf{U} に周期性を仮定すると、 m, n の組み合わせによっては、想定する正弦曲線が冗長になる場合がある. $U_{k,t}$ を最小限の周期関数の重み付き和で表現するために、本発明では α と β に以下に示す group sparsity を考慮する. ここで、係数 α と β を M 個の異なるグループに分け、group sparsity を誘導するた

めの項

$$L(\alpha, \beta) = \sum_{k,m} \left\{ \left(\sum_n \alpha_{m,n}^k \right)^{p/2} + \left(\sum_n \beta_{m,n}^k \right)^{p/2} \right\} \quad (7)$$

を定義する. $L(\alpha, \beta)$ は、高々数個の α と β のグループのみが非ゼロの値を取り、他の全てのグループでは値が 0 になっている場合、小さい値を取る. 従って、 $L(\alpha, \beta)$ を目的関数に含めて最小化することにより、 $U_{k,t}$ を最小限の周期で表現できる. 図 1 は group sparsity を考慮した α と β の推定結果を示している. 数種類の m に関連する α, β のグループでのみ、 α と β の値が大きくなっている. グループ m ごとに、倍音 n 方向の二乗和を最小化することにより、 n 方向の二乗和が 0 に近くなった場合、全ての n で該当する α, β の値が強制的に 0 に変換される. スパース性を保ちつつ、 n 方向では複数の係数が非ゼロになることを許容している.

提案手法では、以上のように各基底スペクトルのアクティベーションに対して、十分な数の周期候補の中から適切に選択された、最小限の周期関数の和で表現される任意の周期を持つ周期的イベントを検出・推定した上で、その周期性を行列分解の枠組みで考慮することが可能になる.

3 最適化アルゴリズム

目的関数である式 (1) を最小化する $\mathbf{H}, \mathbf{U}, \alpha, \beta$ の値を求めるアルゴリズムについて説明する. 繰り返し計算により目的関数を最小化することができる. 各グループでは、それぞれ $\mathbf{H}, \mathbf{U}, \alpha$ と β について最適化を行う 3 段階のステップを実施する. 最適化を解析的に行うことは不可能であるが、補助関数法を用いて、目的関数の各項に対して設定した上界関数 I^+, J^+, L^+ を最小化することにより、間接的に $\mathbf{H}, \mathbf{U}, \alpha, \beta$ の値を最適化する更新式を導出することができる. 上界関数を $\mathbf{H}, \mathbf{U}, \alpha, \beta$ に関して最小化するステップと、目的関数と上界関数の値が等しくなるように補助変数を更新するステップを繰り返すことで、目的関数を単調に降下させることができる.

まず、 $I(\mathbf{H}, \mathbf{U})$ の上界関数 I^+ については、通常の NMF と同様に、Jensen の不等式を用いて以下のように設定することができる.

$$I^+(\mathbf{H}, \mathbf{U}) = \sum_{\omega,t} \left(Y_{\omega,t} \log Y_{\omega,t} - Y_{\omega,t} \sum_k \lambda_{\omega,k,t} \log \frac{H_{\omega,k} U_{k,t}}{\lambda_{\omega,k,t}} - Y_{\omega,t} + X_{\omega,t} \right) \quad (8)$$

補助変数 λ は $\sum_k \lambda_{\omega,k,t} = 1$ を満たす正の重み定数であり、等号成立条件は $\lambda_{\omega,k,t} = \frac{H_{\omega,k} U_{k,t}}{\sum_k H_{\omega,k} U_{k,t}}$ である.

次に、 $J(\mathbf{U}, \alpha, \beta)$ の上界関数 J^+ は以下のように設計できる.

$$J^+(\mathbf{U}, \alpha, \beta) = \sum_{k,t} \left\{ \frac{1}{U_{k,t}} + b(\xi) U_{k,t} + c(\xi) - 2 \log U_{k,t} f_{k,t} + (f_{k,t})^2 \right\}. \quad (9)$$

ここで、 $b(\xi) = \frac{2 \log \xi}{\xi} + \frac{1}{\xi^2}$, $c(\xi) = (\log \xi)^2 - 2 \log \xi - \frac{2}{\xi}$ である. 補助変数 ξ の等号成立条件は $\xi = U_{k,t}$ であ

る。なお、上記の上界関数を設計するにあたって使用した不等式は [3] で用いられている。

最後に、 $f(x) = x^{p/2}$ に接する放物線が $x > 0$ のとき常に $f(x)$ 以上の値を取ることを利用し、 $L(\alpha, \beta)$ に関する上界関数 L^+ を以下のように設計する。

$$L^+(\alpha, \beta) = \sum_{k,m} \left\{ \frac{p}{2} \phi_{k,m}^{p/2-1} \left(\sum_n \alpha_{m,n}^k - \phi_{k,m} \right) + \phi_{k,m}^{p/2} + \frac{p}{2} \nu_{k,m}^{p/2-1} \left(\sum_n \beta_{m,n}^k - \nu_{k,m} \right) + \nu_{k,m}^{p/2} \right\} \quad (10)$$

$\phi_{k,m} > 0$, $\nu_{k,m} > 0$ であり、等号成立条件は $\phi_{k,m} = \sum_n \alpha_{m,n}^k$, $\nu_{k,m} = \sum_n \beta_{m,n}^k$ である。

続いて、上述した上界関数を用いて \mathbf{H} , \mathbf{U} , α , β の更新式を導出する。更新式は、上界関数を偏微分したものを 0 と置いて、各変数について解くことにより得られる。 \mathbf{H} , \mathbf{U} の公式は以下のように導出される。

$$H_{\omega,k} \leftarrow \frac{\sum_t Y_{\omega,t} \lambda_{\omega,k,t}}{\sum_t U_{k,t}}, \quad (11)$$

$$U_{k,t} \leftarrow \frac{d_{k,t} + \sqrt{d_{k,t}^2 + 4w_1 a_{k,t}}}{2 d_{k,t}}, \quad (12)$$

ここで $a_{k,t} = \sum_{\omega} H_{\omega,k} + w_1 b(\xi)$, $d_{k,t} = \sum_{\omega} Y_{\omega,t} \lambda_{\omega,k,t} + w_1 2f_{k,t}$ である。 α , β は、以下のような $\alpha_{m,n}^k$ と $\beta_{m,n}^k$ に関する連立方程式を解くことにより更新できる。

$$w_1 \sum_{m',n'} \left(\sum_t \cos n' P_{m'} t \cos n P_{m'} t \cdot \alpha_{m',n'}^k + \sum_t \sin n' P_{m'} t \cos n P_{m'} t \cdot \beta_{m',n'}^k \right) + w_2 p \phi_{k,m}^{p/2-1} \alpha_{m,n}^k = \sum_t \log U_{k,t} \cos n P_{m'} t \quad (13)$$

$$w_1 \sum_{m',n'} \left(\sum_t \cos n' P_{m'} t \sin n P_{m'} t \cdot \alpha_{m',n'}^k + \sum_t \sin n' P_{m'} t \sin n P_{m'} t \cdot \beta_{m',n'}^k \right) + w_2 p \nu_{k,m}^{p/2-1} \beta_{m,n}^k = \sum_t \log U_{k,t} \sin n P_{m'} t \quad (14)$$

連立方程式 $\mathbf{AX} = \mathbf{B}$ の解は、例えば Matlab では $\mathbf{X} = \mathbf{A} \setminus \mathbf{B}$, Python 言語の NumPy では $\mathbf{X} = \text{numpy.linalg.solve}(\mathbf{A}, \mathbf{B})$, R 言語では $\text{solve}(\mathbf{A}, \mathbf{B})$ とすることにより解くことができる。

4 評価実験

本章では、提案手法を使用した教師なし音源分離、教師あり音源分離の評価実験の結果をそれぞれ示す。音楽データセットとしては、SiSEC2015¹で“Professionally-produced music recordings (MUS)”として公開されている“The Mixing Secret Dataset 100 (MSD100)”を用いた。データセットは異なるジャンルを含む 100 のヴォーカル付き楽曲で構成されている。混合信号に加え、歌手または楽器毎に分離された分離信号も公開されている。

実験データとして、下記の 3 楽曲のデータを用いた。

¹ <https://sisec.inria.fr/professionally-produced-music-recordings/>

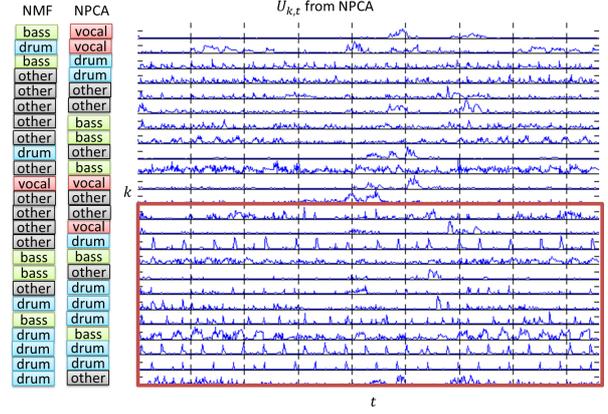


Fig. 2 Estimates of $U_{k,t}$ for musicA1 for unsupervised audio source separation.

- musicA: AM Contra - Heart Peripheral (Electronic Dance Pop)
- musicB: Actions - Devil's Words (Power Pop)
- musicC: Actions - South of the Water (Power Pop)

3 楽曲の分離信号は、全てベース、ドラム、その他の楽器、ヴォーカルの 4 トラックであった。

本実験では、musicA1, musicA2, musicB, musicC として、それぞれ上記楽曲の一部を抽出した。抽出区間は、順番に 1:50-2:00, 0:50-0:60, 1:47-1:57, 1:07-1:17 である。各抽出区間は、大きな旋律の変化を避けて、人手で決定した。音楽データはモノラルでサンプリング周波数は 22.050kHz とした。短時間フーリエ変換におけるフレーム長は 32ms とした。フレームシフト長は 16ms とした。NPCA における \mathbf{H} , \mathbf{U} の初期値には、通常 NMF による分解結果を使用した。NMF における \mathbf{H} , \mathbf{U} の初期値には乱数を使用した。正則化パラメータ w_1 , w_2 はともに 0.2 とし、 $p = 1.0$ とした。

まず、教師なし音源分離において、提案手法 NPCA が分離信号を使用することなく周期的に発生する楽音を抽出できることを検証する。図 2 は musicA1 に対するアクティベーション $U_{k,t}$ の推定結果を示す。基底数 K は 24 とした。図中下部の太枠内の 12 個の基底にのみ、周期性制約を適用した。以下、この 12 基底を周期基底と呼ぶ。 α , β の推定は、周期基底のみで行った。図中で各基底のアクティベーションの左に付けられた楽器ラベルは、分離信号を人手で聴いて付加した。NPCA による分離結果は右側のラベルに示されている。周期的に発音される音と、非周期的に発音される音を分離して、主に周期基底において周期的に発音される音が抽出されていることが確認できる。周期的なリズムが想定されるドラムやベースの音は主に周期基底に、ランダムなリズムが想定されるボーカルは主にそれ以外の基底に割り当てられている。その他の楽器についても周期成分と非周期成分がそれぞれ周期基底、それ以外の基底に分類されていることが分かる。

図 2 の左側の楽器ラベルは、同じデータに対して通常の NMF を適用した場合の結果を示す。多くの基底のアクティベーションが、ドラムやベースの周期的なリズムに影響され、周期的でないその他の楽器など

Table 1 SNR improvement (dB) under each section of music applying NMF and NPCA.

data	model	bass	drum	other	vocal	average
musicA1	NMF	1.59	9.66	11.27	15.02	9.39
	NPCA	2.35	9.61	11.14	14.55	9.41
musicA2	NMF	2.40	10.48	10.89	12.91	9.17
	NPCA	3.19	10.44	10.71	12.79	9.28
musicB	NMF	10.41	15.80	4.45	9.19	9.96
	NPCA	10.56	15.87	4.40	9.17	10.00
musicC	NMF	9.80	18.11	7.02	8.96	10.97
	NPCA	9.77	18.16	6.51	9.20	10.91
average	NMF	6.05	13.51	8.41	11.52	9.87
	NPCA	6.47	13.52	8.19	11.43	9.90

においても、ドラムやベースのリズムに該当するアクティベーションの微小な増加がみられる。アクティベーションの左側に示された楽器ラベルにおいても、多くの基底がベース（11基底）とドラム（8基底）に割り当てられており、ヴォーカルは部分的に1基底に抽出されているのみである。これに対して、提案手法NPCAでは、周期成分と非周期成分が分離され、4基底でヴォーカルの大部分を表現できており、分離性能もNMFと比べて向上していることが確認できた。実験を通して、提案モデルが周期的にアクティベートされる基底を限られた基底数の中にまとめて扱うことができることが示され、この特長は精度と分離信号の解釈性とのトレードオフを解決するのに有効であることが示唆される。加えて、提案モデルを使用すれば、例えば図1のような形で α と β の値を参照することにより、各基底にどのような周期と位相を持つ周期性が推定されたのかを把握することもできる。このようなリズム構造の把握は、自動採譜や楽曲理解のための分析において重要である。

続いて、教師あり音源分離においてNPCAの性能を通常のNMFと比較する。教師あり音源分離の学習フェーズでは、学習データとして分離信号を用いて各基底スペクトルを事前に学習する。分離フェーズでは、テストデータとして混合信号を用いて事前に学習した基底スペクトルを用いて音源分離を行う。この時、基底行列 H の更新は行わず、 U, α, β のみ最適化する。NPCA・NMFともに全ての楽器に対して基底数を6とした。したがって、4楽器による混合信号の分離時の基底数 K は24となる。定量的に分離性能の比較を行うために、signal-to-noise ratio (SNR)を評価指標として用いる。SNRは

$$SNR(dB) = 10 \log_{10} \frac{\sum_{\omega,t} s_{i,\omega,t}^2}{\sum_{\omega,t} (s_{i,\omega,t} - \hat{s}_{i,\omega,t})^2} \quad (15)$$

で定義される。ここで、 $s_{i,\omega,t}$ は入力された楽器 i の分離信号の複素スペクトログラムを、 $\hat{s}_{i,\omega,t} = Y_{\omega,t} \frac{\sum_{k \in i} H_{\omega,k} U_{k,t}}{\sum_k H_{\omega,k} U_{k,t}}$ はWinner filterにより得られる分離信号の複素スペクトログラムを示す。SNRが高いほど、分解性能が高いことを示す。以下、本稿では、SNR improvement, すなわちSNRとSNR-allの差を評価のために用いる。SNR-allは $\hat{s}_{i,\omega,t}$ の代わりに混合信号の複素スペクトログラムを用いて算出したSNRの値である。表1はNMFとNPCAによる音源分離結果から算出したSNR improvementの値の一覧である。結果から、特にベースやドラムなどの周期性が想定される楽器における分離性能の向上が確認できる。musicCにおいては、ヴォーカルパートにおいてもSNR improvementの値が向上している。実

際、musicCのヴォーカルの旋律は抽出した10秒間で2回繰り返されており、提案法が多く周期的なリズムを奏でる楽器に対して分離性能を向上できることが分かる。楽器によってはSNR improvementの値が低下している場合があるものの、周期性という強い制約を与えているにも限らずSNR improvementの平均値がNMFよりも僅かながら増加していることが確認でき、教師あり音源分離の枠組みにおいても提案手法が音源分離性能を向上できることが示された。

5 まとめ

本稿では、NMFの目的関数において、どの音源がいつ発音するかを示すアクティベーションに任意の周期性を仮定する項を付加し、その周期性形状を推定しながら最適化を行うアルゴリズムを導出した。教師あり・教師なし音源分離の両方で提案手法の有効性を評価し、周期的に発生するイベントを抽出することにより音源分離性能が向上することを示した。今後は、特に教師あり音源分離において、各音源における周期性の強さに応じて w_1 の値を自動的に調整するような手法を改良したい。また、提案手法を可変基底NMF[5]と融合させることにより、周期的に基底行列を変化させる手法を考案したい。

参考文献

- [1] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [2] Paris Smaragdis and Judith C Brown, "Non-negative matrix factorization for polyphonic music transcription," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 177–180, 2003.
- [3] Hirokazu Kameoka *et al.*, "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," in *ICASSP 2012*, pp. 5365–5368, 2012.
- [4] Tuomas Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.
- [5] Masahiro Nakano *et al.*, "Nonnegative matrix factorization with markov-chained bases for modeling time-varying patterns in music spectrograms," in *Latent Variable Analysis and Signal Separation*, pp. 149–156. Springer, 2010.
- [6] Kazuki Ochiai *et al.*, "Explicit beat structure modeling for non-negative matrix factorization-based multipitch analysis," in *ICASSP 2012*, pp. 133–136, 2012.
- [7] Rafi Zafar and Bryan Pardo. "Repeating pattern extraction technique (REPET): A simple method for music/voice separation," in *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, no.1, pp. 73–84, 2013.