Vol.2018-MUS-119 No.21
Vol.2018-SLP-122 No.21
2018/6/16

Chroma領域でのNMFとその和声推定への応用

高橋 拓椰[1,†1,a)]　保利 武志[1,†1,b)]　Christoph M.Wilk[1,†1,c)]　嵯峨山 茂樹[1,†1,d)]

概要：本稿では, 音楽音響信号の Chroma 領域における NMF と, 和声推定への応用を検討する. ある単一楽器による学習用データに対する NMF で得られた基底を, 他の学習用データと同一の楽器による演奏データに対する NMF でも共有する. これによって得られたアクティベーションでは, 楽器音の倍音構造に起因する不要成分が抑制されることが期待される. 特に, 和声推定ではピッチクラス情報が重要であるので, 基底を 12 次元に圧縮でき, 計算速度の向上も同時に期待できる. また, これらの手法によって得られたアクティベーションを出力確率と見なし, 統計的に得た和声の遷移確率を用いることで和声進行は HMM でモデル化可能である. 不要成分の少ないアクティベーションと統計データを複合的に HMM のモデルに用いることによって和声推定精度の向上が期待できる. 評価実験として, 倍音成分の異なる楽器の Chroma の本手法による不要成分軽減の実証, また, 和声ラベルが付与されたピアノ楽曲データに対する本手法と従来の Chroma における和声推定の精度を比較することによって, 本手法の和声推定への有用性を示す.

1. はじめに

本稿では, ”教師あり Chroma-NMF”による Chroma 列に対する非負値行列因子分解を行う手法ならびにそれらとマルコフモデルを用いた和声推定手法について提案する.

音楽音響信号を解析する技術は, 和声推定, 調推定, 自動採譜などの場面で活躍する. 藤島は, 楽曲の 12 のピッチクラスをそれぞれ示す Chroma を用いた和声推定手法を提案した [1]. また, Lee らは, Hidden Markov Model(HMM) を用いた音楽音響信号の特徴量分類の手法を提案した [2]. 齋藤らは, 楽曲の調波構造を抑える手法を用いて, その手法によって生成されたスペクトル（調波構造を抑えている）を Chroma 化した Specmurt Chroma を用いた和声推定法を提案している [3],[4]. また, 音楽理論の特性を生かした手法として, 上田らは機能和声を使った転調を考慮した和声推定方法 [5] を, 植村らは, 和声の類似性を示す Doubly Nested Circle of Fifths を用いた和声推定手法 [6] を, Mauch らは, 和声とその構成音の関連性に着目した和声推定法 [7] を, Mauch らは, 音楽の繰り返し構造など楽曲構造を元にダイナミックベイジアンネットワークを用いた和声推定方法を提案している [8],[9]. 他にも, 蔵内らによる, 異なる和声であれば谷となる周波数（スペクトルディップ）が異なることを利用した和声推定法を提案しており [10], さらに黒川らは, このスペクトルディップ法とクロマベクトルを併用し和声解析を行った [11]. NMF を用いた音楽音響信号解析も盛んに行われている. Raczyński らは, 調波構造型の音名に対応する数の基底スペクトルを初期値として音楽音響信号を分析した [12]. 丸尾らは, Bayesian HMM を利用した和声認識と Bayesian NMF を用いた音高推定を組み合わせたコード制約付き NMF によって和声推定を行った [13]. Deng らはニューラルネットワークを使用した和声推定方法を提案している [14],[15]. さらに, Sigtia らは, HMM で行なっていた和声解析部分を recurrent neural network に置き換えた和声解析手法を提案している [16].

このように Chroma による音楽音響信号の解析手法は, 多く使われてきているが, 従来の Chroma では, 調波構造ごと Chroma 化することにより不要成分（倍音成分など）を含んだスペクトログラムが生成される. そこで本稿では, Chroma を NMF におけるアクティベーションであるとみなした時, その不要成分を NMF によって軽減する手法を提案する. 本手法は, NMF の基底を単音の Chroma で構成される教師データによって学習し, その基底を用いて Chroma を NMF することによって, 倍音成分による Chroma の誤差を軽減したアクティベーション出力が期待できる. また, 従来のパワースペクトルに対する NMF による音楽音響信号の解析に比べ基底を 12 次元に圧縮しているため, 計算速度の向上や統計的誤差の軽減が期待できる. また, 本手法を用いた和声解析には HMM を用いるためニューラルネッ

[1] 情報処理学会
IPSJ, Chiyoda, Tokyo 101–0062, Japan
[†1] 現在，明治大学
Presently with Meiji University
[a)] ev60552@meiji.ac.jp
[b)] hori@meiji.ac.jp
[c)] wilk@meiji.ac.jp
[d)] sagayama@meiji.ac.jp

Vol.2018-MUS-119 No.21
Vol.2018-SLP-122 No.21
2018/6/16

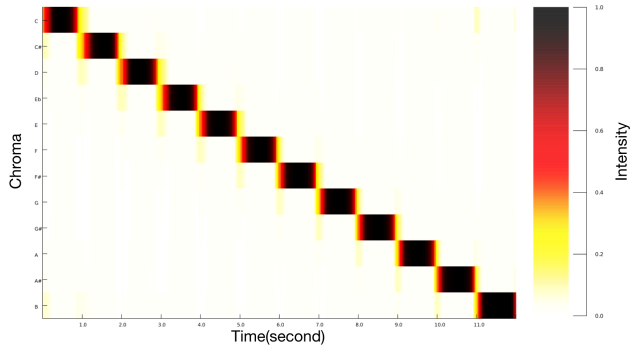

図 1　sin 波による 12 のピッチクラスを演奏した時の Chroma

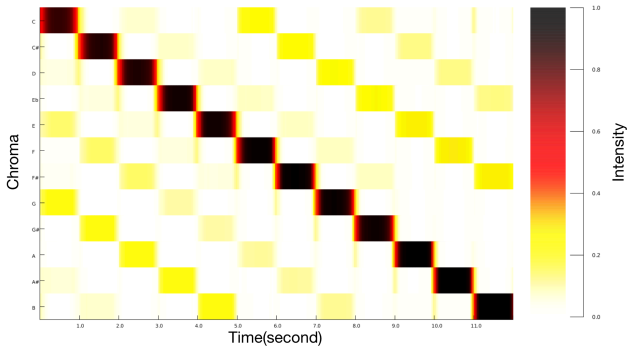

図 2　鋸歯状波による 12 のピッチクラスを演奏した時の Chroma

トワークを用いた手法のように多くのデータを要さない. さらに, 本手法によって調波構造を学習することができれば, ノイズを多く含む音源に対しての和声推定の性能の向上も期待できる.

2. 教師あり Chroma 領域での NMF と和声推定

2.1 Chroma 特徴量

藤島 [1] の提唱した PCP(ここでは Chroma と呼ぶ)ch(k,t) は, 半音幅の Constant-Q Transform (CQT) の絶対値の 2 乗値 $\Phi(f,t)$ から,

$$\mathrm{ch}(k,t) = \sum_{i=0}^{n} \Phi(12i+k,t),\ k = 0 \cdots 11 \qquad (1)$$

によって得ることができる. ただし, n は取得するオクターブの数, $\Phi(f,t)$ は時刻 t の CQT の帯域番号 f のパワーを表わし, ch(k,t) は時刻 t における音名番号 k のパワーを表す. ただし, ここでは, $\sum_{k=0}^{11} \mathrm{ch}(k,t) = 1$ となるように正規化している.

図 1 は, ピッチを半音単位に C から B まで 1 秒ごとに変化させたの正弦波の Chroma である. 図 2 は, 鋸歯状波で同様に得た Chroma である. このように, Chroma ベクトル列は, 実際の楽器音を分析すると単音であったとしても不要成分が倍音構造などに起因して出現する.

2.2 KL ダイバージェンスに基づく非負値行列因子分解

非負値行列因子分解（NMF）は, 一つの非負行列を二つの非負行列で近似する手法である. 応用分野としては, 画像, 音, 生体信号などが挙げられる. 特に, 音の分野では非負の行列となるパワースペクトルを扱うために NMF を適応することができる. 非負の行列（スペクトログラム）$\boldsymbol{Y} \in \mathbb{R}^{M\times N}$ は基底スペクトル $\boldsymbol{W} \in \mathbb{R}^{M\times R}$ とアクティベーション $\boldsymbol{H} \in \mathbb{R}^{R\times N}$ が $\boldsymbol{Y} \approx \boldsymbol{WH}$ となるように分解される. [12] を参考に KL divergence 規準の NMF アルゴリズムを使用する. 行列 $\boldsymbol{X}$ と行列 $\boldsymbol{Y}$ の要素積を $\boldsymbol{X} \odot \boldsymbol{Y}$ とすると, 基底スペクトル $\boldsymbol{W}$ とアクティベーション $\boldsymbol{H}$ は, 更新式

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \odot \frac{\boldsymbol{W^T}\dfrac{\boldsymbol{Y}}{\boldsymbol{WH}}}{\boldsymbol{W^T 1}} \qquad (2)$$

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \odot \frac{\dfrac{\boldsymbol{Y}}{\boldsymbol{WH}}\boldsymbol{H^T}}{\boldsymbol{1H^T}} \qquad (3)$$

を反復することによって求めることができる. ただし, 行列同士の割り算は各要素ごとに行う.

2.3 教師あり Chroma-NMF の定義

我々は, Chroma 自体を NMF によって分解することで, 倍音成分を基底に吸収させ, より不要成分の少ないアクティベーションを推定する手法として “Chroma-NMF” を提唱する. 音楽音響信号の Chroma を NMF するために, 12 のピッチクラスに対応するように基底ベクトルの数を 12 個に設定し, 初期値として各音名に対応する箇所のみを他の音よりも大きい数値を設定する. $\boldsymbol{W}$ の各行の要素は, 各ピッチクラスの Chroma の値を表し, $\boldsymbol{W}$ の列は, 各音名に対応する基底ベクトルを表す. アクティベーションの行は, 各時刻における各基底のパワーを表し, アクティベーションの各列は音名を表す. 解析を行う対象の Chromagram（Chroma ベクトルの時系列）を $Y \in \mathbb{R}^{M\times N}$ を用いて以下の制約に基づいた教師ありの NMF を行う.

- 教師データに基づき, NMF の反復計算を行う際に基底 $\boldsymbol{W}$ のみを更新することで, 基底 W を学習する.
- そこで, 得られた学習済み基底 $\boldsymbol{W}$ を用いてテストデータを NMF する. この時, 基底 $\boldsymbol{W}$ は更新せずアクティベーション $\boldsymbol{H}$ のみを更新する.

ただし, 基底 $\boldsymbol{W}$ を教師データ $Y^{(\mathrm{train})}$ によって学習させる際に, 教師データの楽譜データを元にアクティベーション $\boldsymbol{H}$ に初期値として設定する. $\boldsymbol{H}$ の初期値を決める際, アクティベーション $\boldsymbol{H}$ は, 楽譜上で演奏を指定している箇所の音名の鳴り始めのパワーを 1 とし, 指数関数に従って減

衰していく音減衰モデルを仮定する．$\boldsymbol{H}$ を固定して $\boldsymbol{W}$ のみを式 (2) を使用して更新する．一回更新が終わったら，各音名の基底を基底の音名がベクトルの先頭に来るようにシフトし，そのベクトルの各要素同士の平均を計算し，それを再び基底 $\boldsymbol{W}$ の列に元の形にシフトし直し，$\boldsymbol{W}$ の初期値として設定して，$\boldsymbol{W}$ を更新する．ただし，基底のスケールの任意性を避けるため，$\sum_m W_{m,r} = 1$ となるように正規化を行う．この手順を KL-ダイバージェンスにおける尤度が十分に小さくなるまで更新を続ける．次に，学習を終え倍音構造を学習した基底 $\boldsymbol{W}$ を用いて $Y^{(\mathrm{test})}$ を NMF していく．この際，$\boldsymbol{W}$ は更新せず $\boldsymbol{H}$ のみを式 (1) を使用して更新し，これを繰り返す．

2.4 HMM と Chroma-NMF を用いた和声推定

藤島 [1] では，和声テンプレートベクトルとのユークリッド距離を取ることによって和声推定を行う提案がなされていた．しかし，[1] の実験で使用された和声テンプレートベクトルは，理想的なベクトル (鳴っているべき音名のパワーが 1，それ以外が 0 の 12 次元ベクトル) であり，Chroma における不要成分を考慮していなかった．しかし，Chroma-NMF によって得られたアクティベーションは，不要成分が取り除かれ各音名が明確にどこで鳴っているかを判断することが容易になったことにより，この和声テンプレートベクトルを使った和声推定の精度の向上が期待できる．この時，HMM を使用することによって，多少の非和声音が混ざったとしても正確な和声ラベルを出力することを期待している．ただし，HMM の隠れ状態は和声ラベルである．モデルの限界として，多くの非和声音を含む音楽音響信号であるとその認識精度は期待できない．さらに，非和声音などによって誤った和声ラベルの認識を行うと HMM 部分の処理によって曲全体を通しての和声の認識精度を低下させると考えられる．

ここでは，[1] を参考にしてヒューリスティックに与えた和声のテンプレートベクトルを用いて HMM の出力確率を決定する．各和声のテンプレートベクトルは，12 個の要素を持ち各音名に対応している．一般的にコードのテンプレートベクトル T_i は，

$$T_i = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ \vdots \\ t_{12} \end{pmatrix} \tag{4}$$

と表すことができる．i は和声の種類を表している．この時，和声に対応する各音名の次元に 1，それ以外に 0 を与え全体に摂動を与えたものを $\sum_i T_i = 1$ となるように正規化した．

これを利用して，ある時刻 t における，アクティベーション $H_t = \{h_1, h_2, h_3, \cdots, h_{12}\}$ は各コードテンプレートベクトル T_i に近似できると仮定し，多変量正規分布の平均が T_i であると仮定した時，確率密度を $y_k(t)$ とすると，

$$y_k(t) = \mathcal{N}(H_t | T_i, \Sigma) \tag{5}$$

と表せる．ただし，H_t の各要素は相関を持たないと仮定し，Σ には，各要素に 1 を持つ対角行列を設定している．また，n はベクトルの次元を表している．$y_k(t)$ を HMM の出力確率として考え，最尤推定することによって和声ラベル系列を得る．

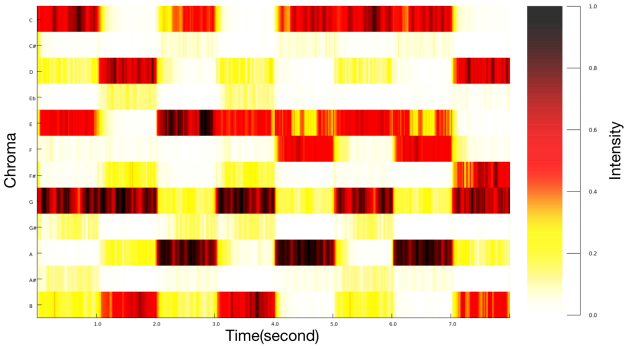

図 3　カノン進行がトランペットによって演奏された時の Chroma

3. 評価実験

3.1 教師あり Chroma-NMF の基底スペクトルとアクティベーション

本項では，教師あり Chroma-NMF を異なった倍音特性を持つ楽器の演奏に対して行なった際の基底スペクトル $\boldsymbol{W}$ とアクティベーション $\boldsymbol{H}$ について評価する．なお，教師データは各楽器の単純演奏を使用する．

3.1.1 実験条件

カノン進行（Cmaj Gmaj Amin Emin Fmaj Cmaj Fmaj Gmaj) の和声を同一テンポ内でかつ同一の音程で演奏され（ベロシティも同じ），1 小節ごとに和声が変化し，ピアノ，トランペット，正弦波，鋸歯状波によって演奏された（倍音成分の違う楽器）4 種類の MIDI 音源を用意する．また，教師データとしてそれぞれ上記の 4 種類の楽器にて演奏され，ピッチが半音単位に C から B まで 1 秒ごとに変化していく MIDI 音源データをを用意した．また，教師データの正解データを音が鳴り始めた時刻ととその音の減衰を考慮して用意した．ただし，教師データ並びにテストデータは，窓シフト 40ms の CQT によってスペクトル解析を行なった．評価の方法としては，Chroma もしくは Chorma-NMF のアクティベーションの曲全体の和音構成音以外の音名の Chroma 成分のパワーの割合を表 1 に示し，比較することで不要成分の減少を評価した．

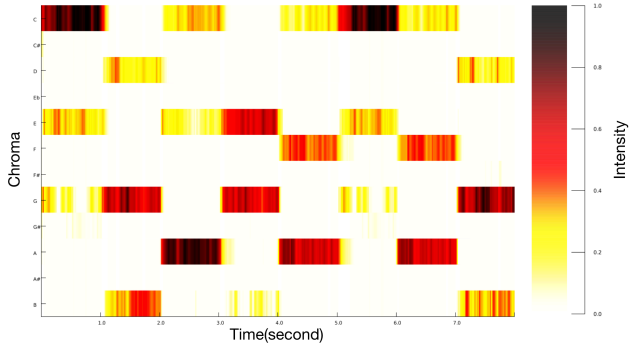

図 4　カノン進行がトランペットによって演奏された時の Chroma-NMF のアクティベーション

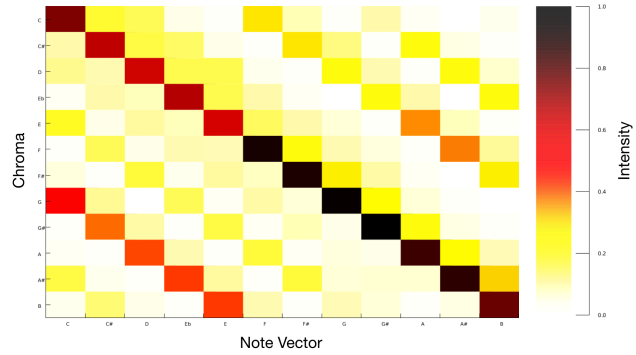

図 5　カノン進行がトランペットによって演奏された時の Chroma-NMF の基底

3.1.2　実験結果

図 4 がトランペットによって演奏されたカノン進行の伴奏を Chroma 化したものを可視化したものであり, 図 5 が同様のものに対して Chroma-NMF を行った時のアクティベーションを可視化したものである. さらに, その時のトランペットの音源を学習済みの基底を図 6 に示す. 表 1 に示すように, 本手法を使用することによって単なる Chroma より, 不要成分のパワーが小さくなっていることがわかった. 本手法によって, 倍音成分を学習した基底 $\boldsymbol{W}$ が test データの NMF の際に, Chroma の倍音成分などを吸収することによって, よりノイズレス (倍音成分などの不要成分を抑えた) なアクティベーションを得ることが期待できる. また, 基底行列の次元が 12 次元に縮小されたことによって計算量を削減できる. さらに Neural Network での手法のように多くのデータからの学習を必要とせず, 同一の倍音特性をもつ演奏の音源と楽譜データがあれば基底を学習可能である. ただし, ある時刻に和声構成音の音名の一部のパワーが Chroma よりも減少することがある. これは, 適切な楽譜データ（減衰などを考慮した）の正解行列をそれぞれの音名のパワーなどを考慮せず与えたことに起因すると考えられる.

表 1　音楽音響信号の Chroma 又は Chroma-NMF による解析結果の不要成分パワーの割合の比較

	Chroma	Chroma-NMF
sin	15.4%	6.9%
sawtooth	32.6%	10.4%
piano	23.8%	7.3%
trumpet	31.6%	4.8%

音楽音響信号は, MIDI 音源の楽器 4 種類で演奏したカノン進行の和声を使用.

3.2　教師あり Chroma-NMF を用いた和声推定

本項では, 教師あり Chroma-NMF を使った音楽の和声推定手法 2.2 についての評価実験について述べる.

3.2.1　実験条件

RWC における楽曲データよりピアノ曲のみをピックアップして教師あり Chroma-NMF のアクティベーションまたは Chroma による和声解析を行った. この時, 著しく和声を捉えることができない曲（単音のみで演奏されている部分が多いなど）は, モデルの限界を考慮し実験の対象から除外した.

実験条件として以下に挙げた事項に基づいて和声推定を行った.

- 窓シフト 40ms の CQT によってスペクトル解析を行なった.
- HMM における遷移確率は, MIREX の和声ラベルデータから統計的に得た.
- 40mS ごとにコードネームを推定した.
- 和声の種類は, “maj”, “min” のみとした.
- 転調を許した.

教師あり Chroma-NMF の基底 $\boldsymbol{W}$ は, 3.1.1 で用いたピアノの教師音源であらかじめ学習した. 本手法との比較実験として [1] を参考に, 得られた Chroma と和声テンプレートベクトルとのユークリッド距離の最小値からの和声推定も行なった. この時, 和声テンプレートベクトルは, 和声内音のパワーを 1 としそれ以外は 0 にセットしたものを用いた. また同様にして, Chroma-NMF と和声テンプレートベクトルとのユークリッド距離の最小値からの和声推定も行なった. 和声テンプレートベクトルは, Chroma のものと同じものを使用した.

評価としては, KSN[17] によって時間ごとにラベルづけされた和声情報を正解データとし本手法によって 40ms ごとに出力された和声を比較して正しい割合を表に示す.

3.2.2　実験結果

表 2 に示すように, ユークリッド距離を用いた Chroma による和声推定よりも, 教師あり Chroma-NMF の Activation と HMM の併用を行なった本手法で得られた方が総合結果としては, 約 1 %認識精度に向上が見られた. また, ユークリッド距離を用いた Chroma での認識よりも Chroma-NMF での認識の方が一部の曲では認識精度の向上が見られた. 一つ一つの曲の認識結果を考察すると, Chroma に

表 2　和声推定実験結果の正解率

	Chroma	Chroma-NMF	Chroma-NMF &HMM
RWC-C22	52.3%	53.0%	54.2%
RWC-C23A	25.5%	17.0%	27.0%
RWC-C23E	40.7%	33.3%	39.9%
RWC-C28	43.1%	47.3%	43.1%
RWC-C29	63.4%	52.1%	62.2%
RWC-C30	63.0%	50.2%	69.9%
RWC-C32	41.0%	41.5%	39.6%
RWC-C33	30.7%	29.2%	30.8%
RWC-C35A	44.2%	43.6%	50.0%
RWC-C35B	36.8%	38.6%	30.3%
RWC-C35C	40.5%	42.0%	39.2%
総合結果	43.4%	41.7%	44.0%

てもともと認識精度が高かった曲のほとんどは, 本手法によってより和声推定精度の向上を示した. ユークリッド距離を用いた Chroma-NMF での認識は, 今回のように和声のテンプレートベクトルとの近似を考えるとき, 非和声音の少ない曲であれば, Chroma による認識に比べより高い和声認識結果を示すことがわかった. さらに, Chroma-NMF と HMM の併用によって多少の非和声音の影響を緩和し, より高い精度の和声推定を可能にした. これは, 和声内音を多く含み, 多少の非和声音が存在する曲に関して, Chroma での認識に比べ, 本手法での認識の方がより高い精度の和声推定を行えることを示している.

しかし, ユークリッド距離を用いた Chroma での和声推定の認識率が低い曲に関しては, 本手法ではかえって認識精度が低くなる傾向にあった. これは, 2.4 で述べたモデルの限界に起因していると考えられる. 実際, 非和声音を多く含む楽曲の認識精度は, Chroma, 本手法ともに低い認識率を示した. さらに, 非和声音を多く含む楽曲では, Chroma よりも本手法の認識精度の方が低くなることが実験によって示された. これは, HMM による多少の非和声音の影響緩和の処理によって, 一度誤った和声認識をしてしまうと, 出力確率の大きな変化が起こる状態まで, 正しいはずの和声内音自体も誤った和声の非和声音だと解釈してしまい認識率を低下させていると考えられる.

さらに, ユークリッド距離を用いた Chroma での認識よりも Chroma-NMF での認識精度の方が低い曲については, 3.1.2 に実験結果で示したように実際にある時刻に鳴っている音名のパワーの一部が減少すると報告したが, これに起因すると考えられる. また, 不要成分を取り除いたアクティベーションと和声テンプレートベクトルとの差を取ったことによって, 非和声音が和声内音であると誤認識する可能性が高まり和声認識精度を低下させたと考えられる. また, 本評価実験は, ピアノ楽曲のみで行なったが解析したい楽曲の一部の楽譜データを入手することができれば, 楽曲の一部の情報を使って基底 $\boldsymbol{W}$ を学習することで他の楽曲も解析することが可能となると考えられる.

非和声音に対応する手法を取り入れることによってより精度の上昇が見込める. 更に, Ueda ら [5] にて, 提案されている手法などと組み合わせることでも, 和声認識精度の上昇を期待することができると考えられる.

4. おわりに

本稿では, 教師あり Chroma-NMF による音楽音響信号の解析手法の提案とこれを用いた和声推定について報告した. Chroma の生成の際に, 調波構造も加算されてしまうことで, 実際の構成音以外のパワーも生起する. この特性を倍音構造を学習した基底を利用した, Chroma-NMF を用いることで構成音以外の音名部分のパワーを抑えることを可能にした. しかし, Chroma-NMF と HMM を用いてある条件下での和声推定への有意性は立証できなかった. 和声推定の結果は全体を通して低いが, これは和声推定結果を向上させる様々な手法を取り入れられていないことが原因としてあげられる. 今後は, 実音源の楽譜データを利用した楽曲の一部で基底を学習することで性能向上につながるかの実験, HMM の和声推定部分の改良, 非和声音を多く含む曲の和声推定方法の検討などを行う予定である.

本研究は科研費基盤A (課題番号 17H00749) の支援を受けた.

参考文献

[1] Takuya Fujishima, “Realtime chord recognition of musical sound: a system using common lisp music,” *Proc. ICMC1999*, pp. 464–467, 1999.

[2] Kyogu Lee and Malcolm Slaney, “Acoustic chord transcription and key extraction from audio using key-dependent hmms trained on synthesized audio,” *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 2, pp. 291–301, 2008.

[3] 齋藤翔一郎, 武山晴登, 西本卓也, 嵯峨山茂樹ほか, “Specmurt 分析と chroma vector を用いた hmm による音楽音響信号の調認識,” *Proc.* 情報処理学会研究報告音楽情報科学 *(MUS)*, Vol. 2005, No. 82 (2005-MUS-061), pp. 85–90, 2005.

[4] Shoichiro Saito, Hirokazu Kameoka, Keigo Takahashi, Takuya Nishimoto, and Shigeki Sagayama, “Specmurt analysis of polyphonic music signals,” *IEEE transactions on audio, speech, and language processing*, Vol. 16, No. 3, pp. 639–650, 2008.

[5] Yushi Ueda, Yuki Uchiyama, Takuya Nishimoto, Nobutaka Ono, and Shigeki Sagayama, “Hmm-based approach for automatic chord detection using refined acoustic features,” In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 5518–5521. IEEE, 2010.

[6] 植村あい子, 甲藤二郎ほか, “Dncof ベクトルとクロマベクトルを用いた和音認識,” *Proc.* 研究報告音楽情報科学 *(MUS)*, Vol. 2011, No. 5, pp. 1–6, 2011.

[7] Matthias Mauch and Simon Dixon, “Approximate note transcription for the improved identification of difficult

chords," In *Proc. ISMIR 2010*, pp. 135–140, 2010.

[8] Matthias Mauch and Simon Dixon, "Simultaneous estimation of chords and musical context from audio," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 6, pp. 1280–1289, 2010.

[9] Matthias Mauch, "Automatic chord transcription from audio using computational models of musical context," 2010.

[10] 蔵内雄貴, 松原正樹, 大野将樹, 斎藤博昭ほか, "周波数スペクトルの谷状点系列による和音推定," *Proc.* 情報処理学会研究報告音楽情報科学 *(MUS)*, Vol. 2008, No. 78 (2008-MUS-076), pp. 125–130, 2008.

[11] 黒川奈桜子, 斎藤博昭ほか, " E-007 スペクトルディップとクロマベクトルの併用による和音推定 (音楽情報科学 (1), e 分野: 自然言語・音声・音楽)," 情報科学技術フォーラム講演論文集, Vol. 11, No. 2, pp. 165–170, 2012.

[12] Stanisław A Raczyński, Nobutaka Ono, and Shigeki Sagayama, "Multipitch analysis with harmonic nonnegative matrix approximation," *Proc. ISMIR 2007, 8th International Conference on Music Information Retrieval*. Citeseer, 2007.

[13] 丸尾智志, 吉井和佳, 糸山克寿, 後藤真孝ほか, "コード制約付き nmf を用いた音高推定に基づくコード認識," 第 *77* 回全国大会講演論文集, Vol. 2015, No. 1, pp. 421–422, 2015.

[14] Junqi Deng, "Large vocabulary automatic chord estimation from audio using deep learning approaches," *HKU Theses Online (HKUTO)*, 2016.

[15] Jun-qi Deng and Yu-Kwong Kwok, "A hybrid gaussian-hmm-deep learning approach for automatic chord estimation with very large vocabulary," *Proc. ISMIR2016*, pp. 812–818, 2016.

[16] Siddharth Sigtia, Nicolas Boulanger-Lewandowski, and Simon Dixon, "Audio chord recognition with a hybrid recurrent neural network," *Proc. ISMIR 2015*, pp. 127–133, 2015.

[17] Kaneko, Hitomi and Kawakami, Daisuke and Sagayama, Shigeki, "Functional harmony annotation database for statistical music analysis," *Proc. the International Society for Music Information Retrieval Conference (ISMIR): Late Breaking session*, 2010.