NMF と基底モデルを用いた多重楽音解析*

☆中鹿亘, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

近年，音楽音響信号から楽音特徴量（音高や発音タイミングなど）の推定を行う楽音解析の研究が盛んになってきている．こうした背景には，WWW の発展とともに音楽データが爆発的に増大し，音楽サービス（音楽検索や音楽推薦など）の要素技術として楽音解析の需要が高まっていることが挙げられる．

単一の旋律だけが続いているような音楽（モノフォニー音楽）の楽音解析は Subharmonic Summation (SHS) [1] や自己相関に基づいた手法 [2] により，比較的高い精度でピッチを推定できた．しかしながら，複数の音が同時刻において混ざり合う音楽（ポリフォニー音楽）の楽音解析は，複数の音源が混在する観測信号からそれぞれの音源情報を推定する問題として定式化され [3, 4]，こうした逆問題を解くことは精度，実時間性という観点から一般的に困難である．

ポリフォニー音楽における多重楽音の推定精度や実時間性の問題を解決する可能性のある手法として，非負値行列因子分解 (Non-negative matrix factorization; NMF) を用いた解析手法が挙げられる [5]．これは，音楽音響信号のスペクトログラムを一つの行列とみなして NMF を実行することで，この行列を音源固有の周波数構造を表す基底行列と，その基底の時間的なゲイン変動を表すアクティビティ行列の積に分解する手法である．これら 2 つの行列を適切に得ることができれば，基底行列から音高 [1] や楽器情報，アクティビティ行列からは発音タイミング，発音長，強度といった楽音特徴量を求めることができる．

しかしながら，これには適切に基底の数を決める必要があり，またそれぞれの基底が解析区間において少なくとも一度は単旋律として出現していなければならないという条件が存在する．基底の数が必要数を超えていれば，例えばピアノ音楽の解析をすると，ピアノのアタック音（ドラムのキックのようなもの）やノイズ音が基底として選ばれてしまったり，逆に基底の数が足りなければ，複数の音源の周波数構造を混合したような基底が表れたりするなど，基底の数が正しく設定されていなければ，音高情報を得ることができない意図しない基底が得られてしまう．これは，NMF が教師なしの行列分解アルゴリズムであり，分解の自由度が高いことに起因していると考えられる．NMF に制約を加えて，ある程度望ましい基底を得ようとする手法 [6] も存在するが，依然として基底数や解析可能条件の問題は解決されていない．

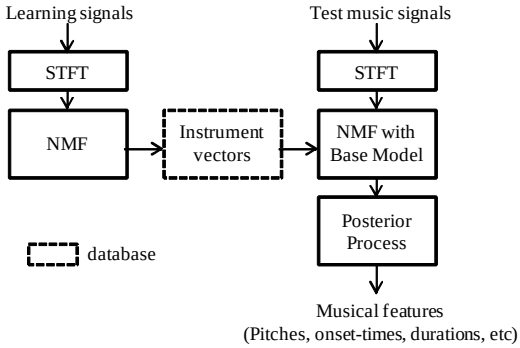

Fig. 1 Flowchart of multiple pitch analysis using NMF with base matrix model.

そこで本研究では，こうした問題点を解決するため，楽器特徴の事前知識（基底モデル）を用いた NMF による多重楽音の解析手法を提案する．すなわち，予め楽器の周波数構造を学習しておき，基底行列を既知として NMF を実行することで，手動で適切な基底の数を設定することなく，多重楽音の解析を行う．基底行列が既知であるため，ノイズなどの意図しない基底が表れるという可能性を考慮する必要がなくなる．また，それぞれの基底モデルに対応するゲインだけがアクティビティ行列の非負値として推定されるので，結果的に解析精度が向上すると期待される．こうした事前知識の利用は，我々人間が音を聞く時，記憶の中にある正解情報とマッチングをとることで，音高や演奏されている楽器を言い当てるという心理作用を倣ったものであり，極めて自然な考え方である．

2 提案手法

提案手法では，楽器特徴の事前知識を用いた NMF によって，ポリフォニー音楽の楽音特徴量を推定する．提案手法の流れ図を Fig. 1 に示す．まず，楽音特徴量の推定を行う前に，基底モデルを学習する (2.1 節)．学習用の音響信号に対しスペクトログラムを求めて，通常の NMF を実行する．このとき，学習に用いる信号は，学習したい基底モデルの音源だけで構成されており，それぞれの音源は単旋律として鳴らされている．音源数が既知なので，このような純粋な信号に対して NMF を実行すれば，理想的な基底行列とアクティビティ行列に極めて近い行列に分解するこ

*Multiple pitch analysis using Non-negative matrix factorization with base matrix modeling. by NAKASHIKA, Toru, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe University)

とができる．この学習ステップで求めた基底行列の列ベクトル集合を，基底モデルとしてデータベースに保持しておく．次に，テストしたい音楽音響信号に対しても同様にスペクトログラムを計算し，予め学習した基底モデルを用いてNMFを実行する．こうして得られたアクティビティ行列に対して，閾値処理を行い，最終的に楽音特徴量（音高，発音タイミング，発音長，強度）を抽出する (2.2 節).

2.1 基底モデルの学習

ある信号のスペクトログラム $\mathbf{V}$ ($\in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}$) に対してNMFを適用すると，

$$\mathbf{V} \approx \mathbf{WH} \tag{1}$$

のように，近似的に基底行列 $\mathbf{W}$ ($\in \mathbb{R}^{\mathbf{m}\times\mathbf{r}}$) とアクティビティ行列 $\mathbf{H}$ ($\in \mathbb{R}^{\mathbf{r}\times\mathbf{n}}$) の積に分解することができる．ここで，$r$ は信号の中に含まれる音源の数に等しい．NMFの計算には，二乗誤差基準により各行列要素の更新を行う．この計算で得られた基底行列の列ベクトル集合 $\mathbf{B} = \{w_l | l = 1, 2, \ldots, r\}$ を，基底モデルとしてデータベースに保存する．本研究では，C2からB5の12音階4オクターブ分 ($r = 48$) のピアノの基底モデルを学習した．

2.2 楽音特徴量の推定

テスト用の音響信号のスペクトログラム $\tilde{\mathbf{V}}$ から，学習した基底モデルによる行列 $\tilde{\mathbf{W}}$ ($\tilde{w}_l \in \mathbf{B}$) を用いて，アクティビティ行列 $\mathbf{H}$ を以下のように繰り返し更新して求めることができる．

$$\mathbf{H}_{ij} \leftarrow \mathbf{H}_{ij} \frac{(\tilde{\mathbf{W}}^T\tilde{\mathbf{V}})_{ij}}{(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\mathbf{H})_{ij}} \tag{2}$$

ここで，$\mathbf{X}_{ij}$ は行列 $\mathbf{X}$ の i, j 成分を表す．式 (2) より求めた $\mathbf{H}$ から，それぞれのノートの発音タイミング，発音長，強度を計算する．閾値処理によってこれらの楽音特徴量を計算するが，具体的な計算手順は紙面の都合上省略する．

3 評価実験

提案手法の有効性を確かめるため，RWCデータベース[1]の音楽データを用いて評価実験を行った．今回用いたデータは，「RWC-MDB-C-2001 No. 43: Sicilienne op.78」と「RWC-MDB-J-2001 No. 9: Crescent Serenade」の2曲である（以後それぞれ#43, #9とする）．#43に対しては16秒，#9は24秒の演奏区間を決めた．それぞれのMIDIデータをピアノ楽器で演奏させ，録音したものをテスト用の音楽音響信号としている．この音響信号に対して，幾つかの手法を用いてMIDIデータに復元した．比較手法としてはスペクトル解析に基づく手法[2]，HTC[3] を用いた．(a) 不正解ノート数/全ノート数, (b) プリファレンススコア (感性基準), (c) プリファレンススコア (再現性基準) の3つの尺度から各手法を比較した．#43と#9の結果を平均したスコアを Table 1 に示す．Table 1 を見れば，いずれの評価基準においても提案手法から最も高いスコアが得られたことが分かる．最後に，提案手法によって#43を解析した例を Fig. 2 に示す．

Table 1 Scores of each method. (%)

	spectro	htc	proposed
(a)	-	61.0	**89.8**
(b)	58.9	61.5	**76.9**
(c)	55.0	56.9	**80.6**

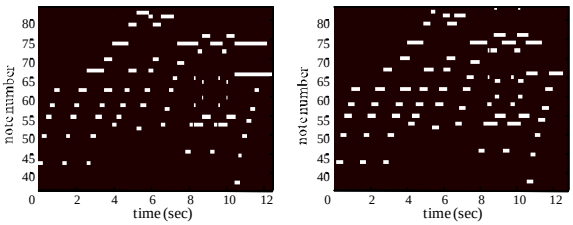

Fig. 2 Original piano-roll of #43 (left), and analysis result using our proposed method (right).

4 おわりに

本研究では，楽器特徴の事前知識を用いたNMFによる音楽音響信号の多重楽音解析手法を提案した．実験結果により，提案手法が主観的にも客観的にも有効であることが示された．

参考文献

[1] D.J. Hermes, “Measurement of pitch by subharmonic summation, ”Journal of ASA, pp.257–264, 1988.

[2] Rabiner, L.R., “On the use of autocorrelation analysis for pitch detection,” IEEE Trans. on ASSP, pp. 24–33, 1977.

[3] H. Kameoka et al., “Harmonic-temporal structured clustering via deterministic annealing EM algorithm for audio feature extraction,” In Proc. ICMIR, pp. 115–122, 2005.

[4] 彦坂健太郎ほか, “可変長セグメントパタンマッチングに基づく楽音の音高・楽器推定,” 日本音響学会春季, pp. 599–600, 2005.

[5] P. Smaragdis and J. C. Brown, “Non-Negative Matrix Factorization for Polyphonic Music Transcription,” IEEE Trans. on WASPAA, pp. 177–180, 2003.

[6] T. Virtanen, “Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria,” IEEE Trans. on ASLP, pp. 1066–1074, 2007.

[1] http://staff.aist.go.jp/m.goto/RWC-MDB/

[2] 採譜の達人, http://www.pluto.dti.ne.jp/ araki/soft/st.html