

確率スペクトルを用いた基底生成モデルと NMF による混合楽音解析*

☆中鹿亘, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

従来の代表的な楽音解析手法として, NMF (非負値行列因子分解) をベースとしたアプローチ [1, 2, 3] が注目を浴びている. これは, 予め大量の音源サンプルを用意しておくことで解析を行う教師あり NMF [3] と, 学習を用いず何らかの制約条件に基づいて解析を行う教師なし NMF [1, 2] に, 大別することができる. しかしながら, 前者では, 可能性のある全ての基底サンプルを用意する必要があるため, 一般にシステムの実用化は困難である. 一方後者のアプローチでは, 機械的に分解しているに過ぎないので意図しない結果が表れる傾向にある. 本研究では, 楽器カテゴリごとに分散を含んだスペクトル包絡を統計的に学習し, その包絡が作り出す基底の組み合わせによって観測信号のスペクトルを表現する手法を提案する. ここで, 楽器ごとに学習するスペクトル包絡を確率スペクトル包絡 (Probabilistic Spectrum Envelope; PSE) と呼び, 周波数-強度平面上の平均曲線と分散曲線で表現されるスペクトル包絡のことを指す. 提案手法ではまず, ガウシアンプロセスを拡張した SPGP+HS [4] を用いて楽器カテゴリごとの確率スペクトル包絡を学習させる. その後教師あり NMF と遺伝アルゴリズムを組み合わせ, 包絡に沿って確率的に生成されるランダム基底集合から, 最適な基底解を探索する. 最後に, 得られたアクティビティ行列から楽音を解析する.

2 提案手法の概要

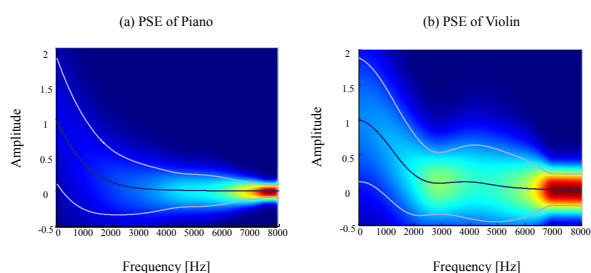


Fig. 1 Examples of probabilistic spectrum envelope; the left is Piano and the right is Violin.

提案手法は, Fig. 2 のように, 確率スペクトル包絡を求める学習ステップと, 実際に楽音解析を行う解析

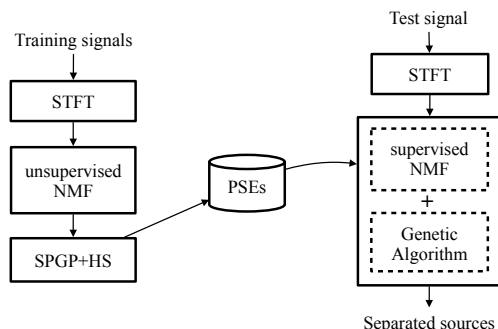


Fig. 2 Modeling of probabilistic spectrum envelopes and analyzing a mixed music signal using them

ステップに分かれる. ここで確率スペクトル包絡とは, 値が一意的である包絡線ではなく, Fig. 1 のような平均と分散で表現されるスペクトル包絡である.

学習ステップでは, まず楽器ごとに用意された学習データ (音階が順に鳴らされた音響信号) の振幅スペクトログラムに対して教師なし NMF を実行する. この NMF によって得られた基底行列から, スペクトルのピーク値 (それぞれの倍音の強度) を取り出し, ガウシアンプロセスをベースとした SPGP+HS [4] により確率スペクトル包絡を学習する. SPGP+HS は, 平均曲線だけでなく分散曲線の推定精度を高めた関数近似手法であり, 平均曲線と分散曲線で表現される確率スペクトル包絡の推定に相応しい.

解析ステップでは, 教師あり NMF と遺伝アルゴリズムを組み合わせる手法によって, テストデータの解析を行う. この遺伝アルゴリズムでは学習ステップで求めた確率スペクトル包絡に基づいた突然変異を起こすように設計されており, 解析対象のスペクトログラムに適合するように, NMF 基底が変形する. 最終的に得られた基底行列, アクティビティ行列を, 最終的な楽音解析の結果とする.

3 学習ステップ

3.1 教師なし NMF による基底スペクトルの抽出

確率スペクトル包絡の学習は, 楽器カテゴリ毎に行われる. 楽器毎に用意される学習データは, その楽器カテゴリに属する音源だけが含まれており音階が順に鳴らされた音響信号である. この信号を短時

* NMF matrix generation using probabilistic spectrum envelope for mixed music analysis. by NAKASHIKA, Toru, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe University)

間フーリエ変換し、得られた振幅スペクトログラム $\mathbf{V} (\in \mathbb{R}^{F \times T})$ に対して、教師なし NMF を用いると、

$$\mathbf{V} \approx \mathbf{W}\mathbf{H} \quad (1)$$

$$\forall i, j, k, \mathbf{W}_{ij} \geq 0, \mathbf{H}_{jk} \geq 0 \quad (2)$$

のように、 $\mathbf{V}$ を 2 つの非負行列の積として表現することができる。ここで、 $\mathbf{W} (\in \mathbb{R}^{F \times R})$ は基底行列、 $\mathbf{H} (\in \mathbb{R}^{R \times T})$ はアクティビティ行列、 R は信号の中に含まれる音源の数である。上に述べたような条件を満たす理想的な信号であれば、教師なし NMF によって得られる基底行列は、純粋な音源のスペクトル集合を表す。

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 [5]。すなわち、式 (2) の元で二乗誤差 $D_{EUC}(\mathbf{V}, \mathbf{W}\mathbf{H}) = (\mathbf{V} - \mathbf{W}\mathbf{H})^2$ を最小化するような $\mathbf{W}$ と $\mathbf{H}$ を求める。各行列要素の更新式は以下のようになる。

$$\mathbf{W}_{ij} \leftarrow \mathbf{W}_{ij} \frac{(\mathbf{V}\mathbf{H}^T)_{ij}}{(\mathbf{W}\mathbf{H}\mathbf{H}^T)_{ij}} \quad (3)$$

$$\mathbf{H}_{jk} \leftarrow \mathbf{H}_{jk} \frac{(\mathbf{W}^T\mathbf{V})_{jk}}{(\mathbf{W}^T\mathbf{W}\mathbf{H})_{jk}} \quad (4)$$

ここで、 $\mathbf{X}_{ij}$ は行列 $\mathbf{X}$ の i, j 成分を表す。式 (3),(4) を繰り返し計算することで、 $\mathbf{W}$ (と $\mathbf{H}$) を求める。

3.2 SPGP+HS による確率スペクトル包絡の推定

スペクトルの包絡線を推定したいので、前節で得られたスペクトルの行列 $\mathbf{W}$ から、まず全てのピーク点 (倍音に相当する周波数とその強度の対) を抽出する。 $\mathbf{W} = [w_1(f) \ w_2(f) \ \cdots \ w_R(f)]$ として、 $r (= 1, \dots, R)$ 番目の音源のスペクトル $w_r(f)$ の基本周波数 f_r を求める。

倍音のインデックスを $h (= 1, \dots, H_r)$ とすれば、 $w_r(f)$ の h 倍音目のピーク点は (f_{hr}, y_{hr}) と表すことができる。ただし、 H_r は $w_r(f)$ のピークの数、 $f_{hr} = h \cdot f_r$ 、 $y_{hr} = w_r(h \cdot f_r)$ である。

$N = \sum_r H_r$ 個のピーク集合 $(\mathbf{f}, \mathbf{y}) = \{(f_n, y_n)\}_n$ を、1次元の SPGP+HS[4] の入力とすれば、平均曲線 μ_f と分散曲線 σ_f を次式のように求めることができる。

$$\mu_f = \mathbf{K}_{ffm} \mathbf{Q}^{-1} \mathbf{K}_{f_m f_n} (\mathbf{\Lambda} + \sigma_\lambda^2 \mathbf{I})^{-1} \mathbf{y} \quad (5)$$

$$\sigma_f = \mathbf{K}_{ff} - \mathbf{K}_{ffm} \mathbf{Q} \mathbf{K}_{f_m f_n} + \sigma_\lambda^2 \quad (6)$$

ただし、 $\mathbf{Q} = \mathbf{K}_{f_m f'_m} + \mathbf{K}_{f_m f_n} (\mathbf{\Lambda} + \sigma_\lambda^2 \mathbf{I})^{-1} \mathbf{K}_{f_n f_m} + \text{diag}(\mathbf{h})$ 、 $\mathbf{Q} = (\mathbf{K}_{f_m f'_m} + \text{diag}(\mathbf{h}))^{-1} - \mathbf{Q}^{-1}$ 、 $\mathbf{\Lambda} = \text{diag}(\mathbf{K}_{f_n f_n} - \mathbf{K}_{f_n f_m} \mathbf{K}_m^{-1} \mathbf{K}_{f_m f_n})$ である。 $\mathbf{K}_{ab}$ はデータ (a, b) 間の、パラメータ θ を持つカーネルの出力値を要素とするグラム行列である。擬似入力 $\bar{\mathbf{f}} = \{\bar{f}_m\}_{m=1}^M$ は入力データ $\mathbf{f}$ のいずれかを表すパラメータであり、

$M \ll N$ を満たす。 $h_m \in \mathbf{h}$ は擬似入力 $\bar{f}_m$ の不確からしさを表すパラメータであり、 $\sigma_\lambda^2, \theta, \bar{\mathbf{f}}$ とともに勾配法によって最適なパラメータを求めることが可能である。SPGP+HS の詳細なアルゴリズムについては紙面の都合上省略するが、詳しくは文献 [4] を参照されたい。

4 解析ステップ

4.1 確率スペクトル包絡に基づくスペクトルのランダム生成

確率スペクトル包絡 $E(f, y; \mu_f, \sigma_f)$ に基づくスペクトル包絡 $e(f)$ は、次式のようにランダムに生成される。

$$e(f) \sim \mathcal{N}(\mu_f, \sigma_f) \quad (7)$$

ここで $\mathcal{N}(\mu, \sigma)$ は平均 μ 、分散 σ の正規分布を表す。

このスペクトル包絡 $e(f)$ に沿った、基本周波数 ν のスペクトル $p(f)$ は

$$p(f) = \max(e(f), 0) \cdot \Psi(f; \nu) \quad (8)$$

と一意に求めることができる。式 (8) で最大値をとっているのは、スペクトルが非負値を取らない制約によるものである。 $\Psi(f; \nu)$ は基本周波数 ν のくし形調波フィルタであり、式 (9) で計算される。

$$\Psi(f; \nu) = \sum_l \exp \left\{ -\frac{(f - \nu \cdot l)^2}{2\lambda_0^2} \right\} \quad (9)$$

ここで l はコンポーネントを示すインデックス、 λ_0 は各コンポーネントの尖度を決定するハイパーパラメータであり、実験的に定められる。

以上の手順で、楽器カテゴリ c 、基本周波数 ν のスペクトルをランダムに生成することができる。

4.2 教師あり NMF による基底の適応度

4.1 節で述べたように、ランダム生成された 1 つのスペクトル包絡から、基本周波数を変えながら複数のスペクトルを作ることができる。さらに楽器カテゴリを変えて得られたスペクトルの集合を $\tilde{\mathbf{W}}$ とする。

一方、解析したいテストデータの振幅スペクトログラムを $\mathbf{X}$ とし、 $\tilde{\mathbf{W}}$ を既知基底行列として教師あり NMF を適用することで、アクティビティ行列 $\tilde{\mathbf{H}}$ を一意に求めることができる。すなわち、式 (3)(4) において $\mathbf{W}$ を更新せず、 $\mathbf{H}$ のみ更新する。

ランダムに生成された基底行列 $\tilde{\mathbf{W}}$ と、得られたアクティビティ行列 $\tilde{\mathbf{H}}$ から、 $\tilde{\mathbf{W}}$ と $\tilde{\mathbf{H}}$ の、 $\mathbf{X}$ に対する適応度 $\Theta(\tilde{\mathbf{W}}, \tilde{\mathbf{H}})$ を次式で計算する。

$$\Theta(\tilde{\mathbf{W}}, \tilde{\mathbf{H}}) = \frac{1}{D_{EUC}(\mathbf{X}, \tilde{\mathbf{W}}\tilde{\mathbf{H}})} \quad (10)$$

Table 1 GA keywords in proposed method.

キーワード	解釈
個体	基底行列 $\tilde{\mathbf{W}}$
個体群	基底行列集合 $\{\tilde{\mathbf{W}}_l\}_{l=1}^L$
遺伝子座	基底ベクトル $\tilde{w}(f)$
適応度	距離 $D_{EUC}(\mathbf{X}, \tilde{\mathbf{W}}\tilde{\mathbf{H}})$ の逆数
交叉	複数の PSE の組み合わせ探索
突然変異	PSE に基づくスペクトル包絡の探索

4.3 遺伝アルゴリズムによる最適基底探索

遺伝アルゴリズムは、遺伝子として表現される複数の個体の中から、適応度の高い個体を選択して交叉・突然変異を繰り返しながら、より適切な解を探索するアルゴリズムである。本研究では、遺伝アルゴリズムを使ってテストデータに適した基底行列を探索することで、精度の高い楽音解析を目指す。表 1 に、提案法における遺伝アルゴリズムの各操作及び単語の解釈を示す。

この解探索法では、まず、 L 個の基底行列を、確率スペクトル包絡に従ってランダムに生成し、式 (10) からそれぞれの適応度を計算する。その後、以下の手順を G 回繰り返す。

1. 前世代の中で最も適応度の高い個体を 1 つ現世代にコピーする。
2. p_{cross} の確率で、基底行列を 2 つ選択して基底ベクトルを交叉させる。
3. p_{mut} の確率で、基底行列を 1 つ選択して基底ベクトルを突然変異させる。
4. 手順 2. と 3. を、現世代の基底行列が L 個になるまで繰り返す。

ここで、基底行列 $\tilde{\mathbf{W}}_l$ を選択する確率を q_l とすると、

$$q_l = \frac{\Theta(\tilde{\mathbf{W}}_l, \tilde{\mathbf{H}}_l)}{\sum_{l=1}^L \Theta(\tilde{\mathbf{W}}_l, \tilde{\mathbf{H}}_l)} \quad (11)$$

と定義する。 p_{cross}, p_{mut} はそれぞれ交叉、突然変異を起こす確率であり、 $p_{cross} + p_{mut} = 1$ を満たす。

交叉は、選択した 2 つの基底行列の各基底ベクトルに対し、0.5 の確率で入れ換える一様交叉を用いる。また、突然変異はそれぞれの基底ベクトルを λ_{mut} (本研究では $\lambda_{mut} = 0.9$) の確率で、基本周波数はそのままに、ランダムにスペクトルを生成したものと入れ替える。すなわち、入れ替えを行う基底ベクトルの調波フィルタは変えないで、確率スペクトル包絡からランダムにスペクトル包絡を生成し、新しくスペクトルを求める。これらの制約により、どの世代のどの個

体の基底ベクトルも、初めに設定した周波数と楽器カテゴリの情報を失うことなく更新される。

遺伝アルゴリズムの結果によって得られた基底行列 $\hat{\mathbf{W}}$ と、そこから計算される $\hat{\mathbf{H}}$ を最終的な楽音解析の結果とする。一度解析が行われれば、これらの行列を用いて様々なタスクに応用することが可能である。例えば、基底行列 $\hat{\mathbf{W}}$ には楽器カテゴリに関する情報 c が含まれているので、 c ごとに対応するアクティビティを出力することで音源分離を実現できる。また、 $\hat{\mathbf{W}}$ には基本周波数、 $\hat{\mathbf{H}}$ にはそれぞれの音源の発音時刻、音価、強度に関する情報が含まれているので、音楽信号の自動採譜に応用することができる。

5 評価実験

5.1 単一楽器の楽音解析

学習ステップで、C1 から B6 の 12 音階 6 オクターブ分 ($R = 72, N = 2705$) の音符が含まれた MIDI データをピアノ音源 (Piano1) で演奏させて録音し、ピアノカテゴリの確率的スペクトル包絡の推定を行った。テストデータは、RWC データベース¹から “RWC-MDB-C-2001 No. 43: Sicilienne op.78” の一部を MIDI 音源で鳴らし、録音したものをを用いている。このとき、様々な環境下で演奏・録音された音響信号を用いて、音源の違いによる頑健性をみる実験を行った。環境条件は (a) Piano1 で演奏、(b) Piano2 で演奏、(c) Piano3 で演奏、(d) 残響レベル 40 で演奏、(e) 残響レベル 100 で演奏となっている。ただし、Piano2 と Piano3 はピアノカテゴリに属する Piano1 とは別の音源、残響レベルは MIDI のリバーブ・レベルを指す。また、遺伝アルゴリズムの個体数 $L = 5$ 、最大世代数 $G = 20$ とした。比較手法として、(1) 教師あり NMF (Piano1 のみ学習) (s-NMF)、(2) 教師なし NMF (us-NMF)、(3) 教師あり NMF (それぞれの環境で録音した基底を学習) (ex. s-NMF) を用いた。それぞれの手法で得られたアクティビティ行列を、適切な閾値で 2 値化し、これを自動採譜の結果とした。

実験結果を図 3 に示す。縦軸は各手法による自動採譜の正解率 $acc[\%]$ を表し、 $acc = \frac{N_{all} - (N_{ins} + N_{del})}{N_{all}}$ 、100 で計算される。ただし、 $N_{all}, N_{ins}, N_{del}$ はそれぞれ全音符数、挿入誤り数、削除誤り数である。上記の 2 値化手法では必ずしも実際の音価と発音継続時間が一致、また、各音源の発音開始時刻が完全に一致することはないので、音価が異なっている、ある許容値 τ だけ発音開始時刻がずれていても正解とみなしている。本研究では $\tau = 0.2$ [sec.] とした。

実験結果から、Piano1 の音源しか学習していない教師あり NMF では、Piano1 で演奏された曲の正解

¹<http://staff.aist.go.jp/m.goto/RWC-MDB/>

率は高いが、Piano2 や Piano3 で演奏，残響をのせると，正解率が低下することが分かる．テストデータに使用されている音源を学習する NMF では，その音源を知っているため，どの環境下でも正解率が比較的高い．提案手法では，Piano1 の音源のみを使用して学習しているにも関わらず，他の環境下で演奏された信号に対しても，比較的高い正解率を示した．このことから，提案手法は未知の楽器に対して頑健であると言える．また，教師なし NMF の結果でも，環境によって正解率があまり変わらないが，教師なしであることから意図しない基底が表れ，その結果最も正解率が低くなったと考えられる．

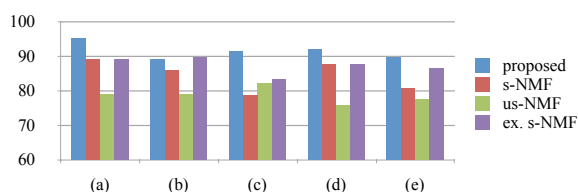


Fig. 3 Accuracy rates of each method.

5.2 複数楽器の楽音解析

複数楽器を用いた実験では，ピアノ音源に加えて，ヴァイオリンの音源を学習させた．テストデータには単一楽器の実験と同じ曲を，図 4 (c) のようにヴァイオリン音源 (紫)，ピアノ音源 (赤) で MIDI を演奏させ，録音したものを用いている．

提案手法による解析結果例を図 4 に表す．同図 (a)，(b) はそれぞれ，遺伝アルゴリズムの最大世代数 $G = 500$ としたときの，初代時，最終世代時の解析結果である．初代では，本来ヴァイオリン音であるはずの幾つかの音源が，誤ってピアノ音源であると推定されているが，世代を交代する度に徐々にテストデータに適応していき，500 世代目では，誤って推定された楽器ラベルがほぼ正されているのが分かる．

従来の教師なし NMF による楽音解析 [1, 2] では，テストデータの音源が未知であっても楽音の解析が可能であったが，推定した基底の音源を特定することができなかった．しかし，提案手法では楽音の解析を行うと同時に楽器の分離ができ，この性質は提案手法の大きな強みであると考えている．

6 おわりに

本研究では，楽器カテゴリごとに分散を含んだスペクトル包絡を，拡張ガウシアンプロセス (SPGP+HS) によって学習しておき，教師あり NMF と遺伝アルゴリズムを組み合わせた確率的なアプローチによる新

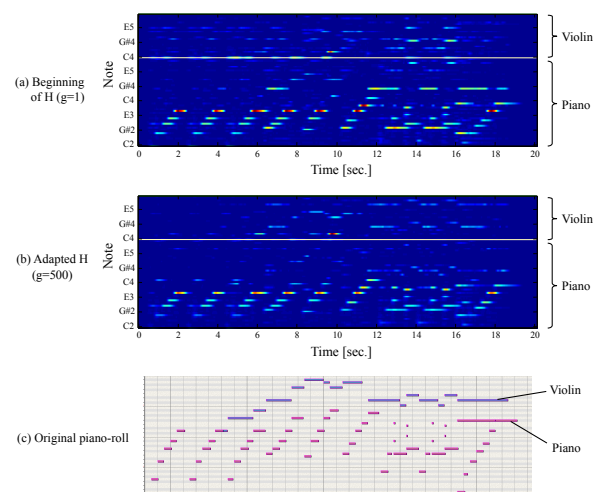


Fig. 4 Results of the experiment with multiple instruments.

しい楽音解析手法を提案した．実験結果により，提案手法が従来手法と比較して未知の楽器に頑健であり，複数楽器混在でも解析を行うことができることを確認した．

参考文献

- [1] P. Smaragdis and J. C. Brown, “Non-negative matrix factorization for polyphonic music transcription,” in *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 177–180, 2003.
- [2] T. Virtanen, “Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 3, pp. 1066–1074, 2007.
- [3] A. Cont, S. Dubnov, and D. Wessel, “Realtime multiple-pitch and multiple-instrument recognition for music signals using sparse non-negative constraints,” in *Proceedings of Digital Audio Effects Conference (DAFx)*, pp. 10–12, 2007.
- [4] E. Snelson and Z. Ghahramani, “Variable noise and dimensionality reduction for sparse Gaussian processes,” in *Proceedings of the 22nd International Conference on Uncertainty in Artificial Intelligence*, Citeseer, 2006.
- [5] D. Lee and H. Seung, “Algorithms for non-negative matrix factorization,” *Advances in neural information processing systems*, vol. 13, 2001.