Vol.2011-MUS-89 No.18
2011/2/12

確率スペクトル包絡に基づく NMF 基底生成モデルを用いた混合楽音解析

中鹿 亘[†1] 滝口哲也[†2] 有木康雄[†2]

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 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 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ので，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共通なスペクトル包絡（確率スペクトル包絡）を統計的に学習し，確率スペクトル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．提案手法ではまず，ガウシアンプロセスをベースとした手法により，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て，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．実験結果から，提案手法が学習データには含まれない未知の音源に対しても頑健であると同時に，複数の音源が混ざっていても解析が可能であることを確かめた．

NMF Matrix Generation Using Probabilistic Spectrum Envelope for Mixed Music Analysis

Toru Nakashika,[†1] Tetsuya Takiguchi[†2] and Yasuo Ariki[†2]

NMF (Non-negative Matrix Factorization) based approaches are garnering much attention in musical signal analysis in recent years. These are roughly classified into two approaches: exemplar-based NMF, in which a large number of samples are used for analyzing a signal, and unsupervised NMF, in which signals are analyzed in some constrains without learning any samples beforehand. However, because the former methods require all the possible samples for the analysis, it is hard to build the practical system of the method. The latter approach should cause unintended results because the method is based on mathematical analysis not perceptual coding. In this paper, we propose a novel method of signal analysis by combining NMF and a probabilistic approach. At the beginning, a common spectram envelope to an instrument, called a probabilistic spectrum envelope (PSE), is learned for each categories using a Gaussian-Process-based approach. On the analyzing stage, basis vectors of NMF are randomly generated from the PSE, and the most befitting vectors can be found by combination of unsupervised NMF and Genetic Algorithm. The experimental results indicated that the method is robust against unknown sound sources, and can properly analyze the signals including multiple sources.

1. はじめに

近年 WWW の発展と共に音楽データが爆発的に増大し，それに伴って音楽信号を対象とした信号処理（音楽情報処理）に関する関心が高まっている．中でも，自動採譜システムや音楽検索など，様々な音楽アプリケーションに応用が可能であることから，音楽音響信号から個々の基本周波数（音階）を推定する楽音解析の研究が特に注目を浴びている．

モノフォニー音楽のように，単旋律が続く音楽の解析では，Subharmonic Summation (SHS)[1] や自己相関に基づいた手法[2] により，比較的高い精度で基本周波数を推定することができた．しかしながら，和音を含む音楽やポリフォニー音楽のように，同時に複数の音階の音が鳴っている信号や，さらに複数の楽器が混ざっている音響信号から個々の音源をシングルチャンネルで推定することは，解析精度や実時間性の観点か

†1 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University
†2 神戸大学自然科学系先端融合研究環
Organization of Advanced Science and Technology, Kobe University

ら，楽音推定の中でも最も困難な問題とされている．本研究では，複数楽器を含むシングルチャンネルのポリフォニー音楽信号から，個々の楽音イベント（音階，強度，発音時刻，音価，楽器ラベル）を推定する楽音解析を対象としている．

こうしたシングルチャンネルにおける多重奏，複数楽器混在の混合楽音解析を実現するため，観測パワースペクトルを拘束付き GMM でモデル化した音モデルの加重混合とみなして最適解を解く手法[3]，観測スペクトルを不特定楽器存在確率と条件付き楽器存在確率の積で表現する手法[4] など，様々な手法がこれまでに提案されてきた．中でも最も有力とされている解析手法の１つとして，非負値行列因子分解 (Non-negative matrix factorization; NMF) を用いた手法がある[5)–8)]．これは，音響信号の振幅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情報（スペクトル）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．得られた２つの行列から，それぞれ音階情報，時間情報（発音時刻，音価，強度）を求めることができ，これにより楽音イベントが推定可能である．

NMF を用いたシングルチャンネル楽音解析は，大別して教師なしのアプローチ，教師ありのアプローチに分けることができる．前者の教師なしアプローチ[5),6)] では，音源の構造を仮定せずに，拘束条件のみを用いて機械的に基底行列とアクティビティ行列の分解を行う．そのため，本来意図しない基底やアクティビティが表れてしまい，解析精度の低下に繋がる．

一方教師あり NMF を用いた手法では，イベント（特定楽器の特定音階）ごとの音響信号からそれぞれのスペクトルテンプレートを学習しておき，そのテンプレートに基づいて解析を行う[7),8)]．こうした教師ありのアプローチでは，比較的高速かつ高精度な結果が得られている．しかしながら，これから解析を行う信号の中に，未学習のイベントが含まれていれば解析精度が落ちてしまうという問題点がある．解析精度を高めるためには，非常に膨大なテンプレートを用意し学習させなければならないが，可能な限りの全てのイベントを収集するのは現実的に極めて困難である．

そこで本研究では，全てのイベントのスペクトルテンプレートを用意するのではなく，特定のイベントを集約したカテゴリ（楽器カテゴリ）ごとに確率的なスペクトルのテンプレートを学習しておき，このテンプレートに従ったスペクトルをランダムに生成することで，未知イベントに対応した教師あり NMF による信号解析手法を提案する．ここで，確率的なテンプレートのこと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．この平均曲線と分散曲線は，楽器カテゴリによって特徴付けられたもの（すなわち楽器カテゴリによって唯一定まるもの）である．提案手法は，楽器カテゴリごとの確率スペクトル包絡を求める学習ステップ，確率スペクトル包絡に基づいてランダムに基底集合を生成し，テストデータの楽音を解析する解析ステップの２つに分けることができる．本研究では，確率スペクトル包絡の学習に，ガウシアンプロセスを拡張した SPGP+HS[9)] と教師なし NMF を，テストデータの解析に，遺伝アルゴリズムと教師あり NMF を用いる．

2. 提案手法の概要

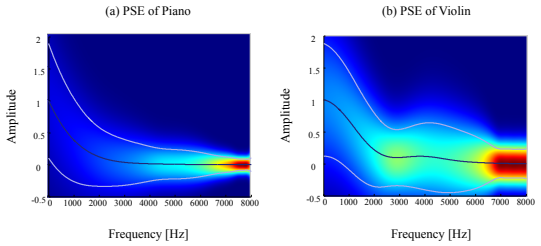

図 1 確率スペクトル包絡の例．左がピアノ，右がヴァイオリンの確率スペクトル包絡である．赤色は確率値が高く，青色は確率値が小さいことを示している．黒線は平均曲線，その上下の白線は平均曲線 ± 分散曲線をプロットしたもの．

Fig. 1 Examples of probabilistic spectrum envelope; the left is Piano and the right is Violin. Red and blue color in the figure indicate the large and small value of probability, respectively. Black line is mean contour, and white lines are mean contour plus and minus variance contour.

2.1 確率スペクトル包絡

楽音解析，楽器音分離に関する従来研究の多くは，「楽器音の各倍音の強度比率（調波構造）は音高によらず一定である」と仮定して楽器音を特徴付けていた[3),10)]．しかしながら実際の楽器音は，音高によって少なからず調波構造が異なっており，この仮定が推定精度を低下させる要因となる．そこで本研究では，楽器音をより適切に特徴付けるため，「１つの楽器カテゴリには，１つの確率的なスペクトル包絡（確率スペクトル包絡）が存在し，楽器カテゴリに属する楽器は全て，その共通の確率的スペクトル包絡を１つ持つ」と仮定する．

確率スペクトル包絡は，図１に示すように，平均曲線と分散曲線で表される確率的なスペクトル包絡である．包絡線の導入により，楽器音の調波構造は音高によって形を変えることができる．このとき，ある音高に対する調波構造は一定ではなく，分散値によって変化する．この揺らぎが楽器を識別する特徴の１つであると考える．また，この仮定により，“Piano” 楽器カ

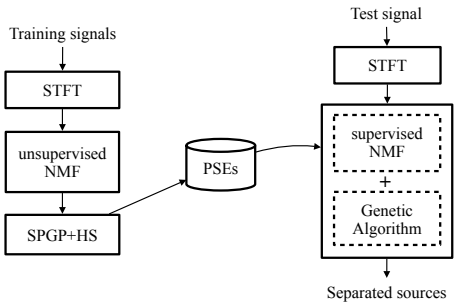

図 2 提案手法のフローチャート.
Fig. 2 Modeling of probabilistic spectrum envelopes and analyzing a mixed music signal using them

テゴリに属する，ある楽器 (Piano1) だけを用いて確率スペクトル包絡を学習すれば，“Piano” カテゴリに属する別の楽器 (Piano2 や Piano3 など) の全ての確率スペクトル包絡を表現できることを示唆している．教師あり NMF による楽音解析では，同じ楽器カテゴリでも調波構造が異なると推定精度が落ちてしまうという問題があったが，提案手法では前述の仮定から，学習時には含まれない未知の楽器のスペクトルを表現できることが期待される．

2.2 提案手法の流れ

提案手法では，楽器のカテゴリごとに分散を含めたスペクトル包絡を学習し，それを基にランダムにスペクトルを生成させ，確率的なアプローチにより複数楽器混在の音楽信号を解析する．提案手法のフローチャートを図 2 に示す．

提案手法は，確率的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ステップに分かれる．学習ステップでは，予め幾つかの学習データを用意する．ここで学習データは，楽器カテゴリごとに用意され，それぞれの音階が順に鳴らされた（同時には演奏されない）音響信号を用いる．次に，学習データ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．音源数を既知として，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とができる．ここで，理想的な基底行列とは，それぞれのイベントのスペクトルを列要素とする行列を意味する．学習に用いる全ての音源には基本周波数が存在すると仮定しているので，そのスペクトルは倍音構造を持つ．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[9] により確率スペクトル包絡を楽器ごとに学習する．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た手法によって，テストデータの解析を行う．具体的には，基底行列とアクティビティ行列の積と，観測スペクトログラムとの距離を評価関数とした遺伝アルゴリズムを用いて，距離が最小となるように基底ベクトルの選択・交叉・突然変異を繰り返し，最適な基底行列，アクティビティ行列を求める．ここで，基底ベクトル（行列）の突然変異とは，学習ステップで求めた確率スペクトル包絡を基にしてランダムにスペクトル基底が生成される現象である．この突然変異と，交叉・選択を繰り返す遺伝アルゴリズムは，確率スペクトル包絡というソフトな解空間から様々な可能性を探索し，テストデータに最も適応した基底行列を効率よく見つけることに等しい．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．学習に用いる音響信号は，次の条件を満たす．

- 楽器カテゴリに属する音源だけが含まれている
- それぞれの音源は同時に鳴らされていない
- 音源の数は既知である

あるカテゴリに属する信号を，サンプリング周波数 f_z で短時間フーリエ変換する．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，

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

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

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[11]．すなわち，式 (2) の元で二乗誤差 $D_{EUC}(\mathbf{V},\mathbf{WH}) = (\mathbf{V}-\mathbf{WH})^2$ 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．

$$\mathbf{W}_{ij} \leftarrow \mathbf{W}_{ij}\frac{(\mathbf{VH}^T)_{ij}}{(\mathbf{WHH}^T)_{ij}} \tag{3}$$

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

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，$r\ (=1,\cdots,R)$ 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．基本周波数は，ゼロクロス法や自己相関法[12] などを用いて計算することができる．倍音のインデックスを $h\ (=1,\cdots,H_r)$ 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 $(\boldsymbol{f}, \boldsymbol{y}) = \{(f_{hr}, y_{hr})\}_{h,r} = \{(f_n, y_n)\}_n$ を，1 次元の SPGP+HS[9] に入力すれば，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．

$$\mu_f = \mathbf{K}_{ff_m}\mathbf{Q}^{-1}\mathbf{K}_{f_mf_n}(\mathbf{\Lambda} + \sigma_\lambda^2\mathbf{I})^{-1}\boldsymbol{y} \tag{5}$$

$$\sigma_f = \mathbf{K}_{ff} - \mathbf{K}_{ff_m}\acute{\mathbf{Q}}\mathbf{K}_{f_mf} + \sigma_\lambda^2 \tag{6}$$

ただし，$\mathbf{Q} = \mathbf{K}_{f_mf'_m} + \mathbf{K}_{f_mf_n}(\mathbf{\Lambda}+\sigma_\lambda^2\mathbf{I})^{-1}\mathbf{K}_{f_nf_m} + diag(\boldsymbol{h})$，$\acute{\mathbf{Q}} = (\mathbf{K}_{f_mf'_m} + diag(\boldsymbol{h}))^{-1} - \mathbf{Q}^{-1}$，$\mathbf{\Lambda} = diag(\mathbf{K}_{f_nf_n} - \mathbf{K}_{f_nf_m}\mathbf{K}_m^{-1}\mathbf{K}_{f_mf_n})$ 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 $\bar{\boldsymbol{f}} = \{\bar{f}_m\}_{m=1}^M$ は入力データ $\boldsymbol{f}$ のいずれかを表すパラメータであり，$M \ll N$ を満たす．$h_m \in \boldsymbol{h}$ 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，$\sigma_\lambda^2, \theta, \bar{\boldsymbol{f}}$ 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 9) を参照されたい．

以上の手順を楽器カテゴリ毎に行う．式 (5),(6) では $\mu(f)$ と $\sigma(f)$ を一般化して表記しているが，楽器カテゴリ c の平均曲線 μ_f^c と分散曲線 σ_f^c を持つ確率スペクトル包絡 $E^c(f;\mu_f^c,\sigma_f^c)$ をデータベースに保存する．ただし $c = 1\cdots C$ であり，C は楽器カテゴリの数である．

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．

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

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

$$p(f) = \max\big(e(f), 0\big) \cdot \Psi(f;\nu) \tag{8}$$

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．

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

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．本研究ではアクティビティ行列の算出に，次で示されるような擬似逆行列を用いた計算法を使用する．

1. $\tilde{\mathbf{H}} = (\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}^T\mathbf{X}$ を計算する．
2. $\tilde{\mathbf{H}} \to \tilde{\mathbf{H}} \in \mathbf{R}_+ = \{x, x \in [0,\infty)\}$ として $\tilde{\mathbf{H}}$ を非負空間へ射影する．
3. $\frac{\tilde{\mathbf{H}}_{jk}}{\|\tilde{\mathbf{H}}\|^2} \leftarrow \tilde{\mathbf{H}}_{jk}$ と正規化する．

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．

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

もし $\tilde{\mathbf{W}}$ が観測データをうまく表現できない，外れた基底行列であれば，上記手順 1. より計算された行列 $\tilde{\mathbf{H}}$ の要素の多くは負値となる．このため，手順 2. で $\tilde{\mathbf{H}}$ を非負空間へ射影したときに失われる情報が多くなる．よって，このような解候補 $(\tilde{\mathbf{W}},\tilde{\mathbf{H}})$ の適応度が必然的に低くなる．一方，$\tilde{\mathbf{W}}$ がテストデータに使用されている音源集合に近い ‘良い’ 基底行列なら，$\tilde{\mathbf{H}}$ も正しく推定され，距離 $D_{EUC}(\mathbf{X},\tilde{\mathbf{W}}\tilde{\mathbf{H}})$ が小さくなるので適応度は高くなる．したがって，式 (10) で表される適応度を，ランダムで生成した $\tilde{\mathbf{W}}$ の良さを表す評価関数として用いることができる．

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．

表 1 提案法における遺伝アルゴリズムの解釈
Table 1 The meanings of keywords of Genetic Algorithm in the 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	複数のスペクトル包絡の組み合わせ探索
突然変異	確率スペクトル包絡に基づくスペクトル包絡の探索

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，

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

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．本研究では，$p_{cross} = 0.9, p_{mut} = 0.1$ と設定する．

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．

各楽器カテゴリが持つ確率スペクトル包絡は，値がはっきりと定まるスペクトル包絡ではないが，遺伝アルゴリズムを用いて確率スペクトル包絡が張る解空間を探索することができる．言い換えれば，楽器カテゴリの確率的にモデル化した特徴量を，確率的な状態を保持したまま楽音解析問題に当てはめて，最適な解を探索することと同等であり，情報量（特徴量）の損失低下という観点から，確率スペクトル包絡による楽器特徴のモデル化と，遺伝アルゴリズムによる楽音解析は非常に相性が良いと言える．

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，$acc = \frac{N_{all} - (N_{ins} + N_{del})}{N_{all}} \cdot 100$ 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，音価が異なっていても，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

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

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．

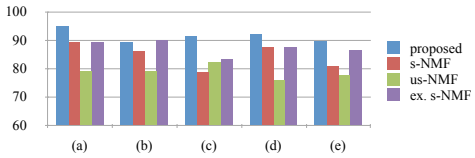

図 3 単一楽器実験の結果．
Fig. 3 Accuracy rates of each method.

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[5),6)] 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り楽音解析を行う，新しい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．

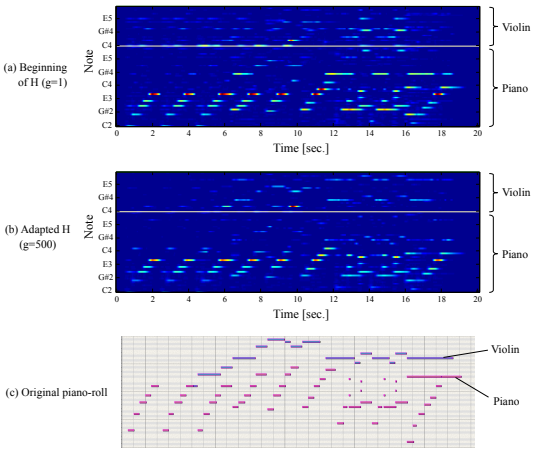

図 4 複数楽器を用いた実験結果．
Fig. 4 Results of the case with multiple instruments.

参 考 文 献

1) Hermes, D.: *Journal of the Acoustical Society of America*, Vol.83, No.1, pp.257–264 (1988).
2) Rabiner, L.: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, Vol.25, No.1, pp.24–33 (2003).
3) Miyamoto, K., Kameoka, H., Nishimoto, T., Ono, N. and Sagayama, S.: *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, IEEE, pp. 113–116 (2008).
4) Kitahara, T., Goto, M., Komatani, K., Ogata, T. and Okuno, H.: *Information and Media Technologies*, Vol.2, No.1, pp.279–291 (2007).
5) Smaragdis, P. and Brown, J. C.: *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp.177–180 (2003).
6) Virtanen, T.: *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 15, No.3, pp.1066–1074 (2007).
7) Schmidt, M. N. and Olsson, R. K.: Single-channel speech separation using sparse non-negative matrix factorization, *In International Conference on Spoken Language Processing (INTERSPEECH)* (2006).
8) Cont, A., Dubnov, S. and Wessel, D.: *Proceedings of Digital Audio Effects Conference (DAFx)*, pp.10–12 (2007).
9) Snelson, E. and Ghahramani, Z.: *Proceedings of the 22nd International Conference on Uncertainty in Artificial Intelligence*, Citeseer (2006).
10) Saito, S., Kameoka, H., Takahashi, K., Nishimoto, T. and Sagayama, S.: *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol.16, No.3, pp.639–650 (2008).
11) Lee, D. and Seung, H.: *Advances in neural information processing systems*, Vol.13 (2001).
12) Huang, X., Acero, A. and Hon, H.: Prentice Hall PTR Upper Saddle River, NJ, USA (2001).