社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

確率スペクトル包絡を用いた混合音解析における制約付きスペクトル生成法の検討

中鹿　　亘†　滝口　哲也††　有木　康雄††

† 神戸大学大学院システム情報学研究科
〒 657-8501 兵庫県神戸市灘区六甲台町 1-1
†† 神戸大学自然科学系先端融合研究環
〒 657-8501 兵庫県神戸市灘区六甲台町 1-1
E-mail: †nakashika@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

あらまし　従来の代表的な信号解析手法の中に，NMF（非負値行列因子分解）を用いたものがある。特に，事例ベースの NMF が音源分離や信号解析の分野において，解析精度・速度といった観点から注目を浴びている．しかしこうした手法は，可能性のある全ての事例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。これまでの我々の研究では，この問題点を解決するため，確率的に生成されるスペクトルを用いて信号を解析する確率スペクトル包絡による手法を提案してきた。しかしながら，この方法では高いスペクトル生成自由度により分離最適解を得ることが困難であった。そこで本研究では，アクティビティ行列要素のスパース性と密集性に着目した新たな制約項を加えることにより，より最適な解に導く信号解析手法を提案する。

キーワード　信号解析，音源分離，教師あり NMF，確率的スペクトル包絡，ガウシアンプロセス，スパース制約

Constrained Spectrum Generation for Mixed Sound Analysis Based on Probabilistic Spectrum Envelope

Toru NAKASHIKA†, Tetsuya TAKIGUCHI††, and Yasuo ARIKI††

† Graduate School of Engineering, Kobe University
Rokkodaicho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan
†† Organization of Advanced Science and Technology, Kobe University
Rokkodaicho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan
E-mail: †nakashika@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

Abstract　NMF (Non-negative matrix factorization) has been one of the most widely-used techniques for signal analysis in recent years. In particular, the supervised type of NMF is garnering much attention in source separation or signal analysis with respect to the analysis accuracy and speed. Because such methods require all the possible samples for the analysis, it is hard to build a practical analysis system. To analyze signals properly even when short of samples, we proposed a probabilistic approach called PSE (probabilistic spectrum envelope) so far, in which spectrum envelopes belonging to an auditory category are randomly generated, and the spectrum is used as a part of supervised basis matrix of NMF. However, this method has a difficulty in obtaining the optimum solution due to a lot of flexibility. In this paper, we propose a new PSE method with sparseness and density constraints which efficiently lead to the more appropriate solution.

Key words　signal analysis, source separation, supervised-NMF, probabilistic spectrum envelope, Gaussian process, sparseness

1. はじめに

機械との自然な対話コミュニケーションを実現するため，複数の話者が同時に発話した混合音からそれぞれの話者の信号を推定（音源分離）したり，雑音を抑圧して音声を強調（雑音除去）するなど，コンピュータが音を聞き分ける技術というものが必要とされている．しかしながら，カクテルパーティ問題として知られているように，複数の音源が混ざり合った音響信号から，個々の音源を推定する逆問題を解くことは一般的に困難である．中でも，シングルチャンネルで録音された信号を用いて音源分離やノイズ除去を行うことは難しい問題とされている．一方，アレー信号処理 [1] や二段 BSS 法 [2] など，マイクを複数用いた手法では解析精度・解析速度の面で成功を収めているが，製造コストの削減や省スペース化の面で，シングルチャンネルのみを用いて信号を解析する方が望ましい．

こうしたシングルチャンネルにおける音源分離を実現するため，factorial HMM を用いた手法 [3]，独立成分分析 (ICA) を用いた手法 [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]〜[10]．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．このようにコンパクト表現された基底行列の情報を基にして，音源分離を実現する．

NMF を用いたシングルチャンネル音源分離は，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]〜[7] では，音源の構造を仮定せずに，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，結果分離を正しく行うことができない．

一方教師あり NMF を用いたシングルチャンネル音源分離では，イベント（特定話者の特定の音素や楽音単位）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分離を行う [8]〜[10]．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，これから分離を行う信号の中に，未学習のイベントが含まれていれば分離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．

そこで我々はこれまでの先行研究として，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（非特定話者の音素や，楽器など）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，未知のイベントを頑健に解析できる確率スペクトル包絡 (PSE; probabilistic spectrum envelope) 法を提案してきた [11], [12]．PSE はスペクトル包絡の平均と分散で表現され，カテゴリ内のスペクトル変動を考慮したイベント不変のスペクトル包絡である．この包絡に従ってランダムに基底を生成すれば，学習時には含まれない未知のイベントのスペクトルを表現することができるため，このスペクトルを列ベクトルとした基底行列を与えれば，教師あり NMF の枠組みで未知イベントに頑健な信号解析が可能となる．

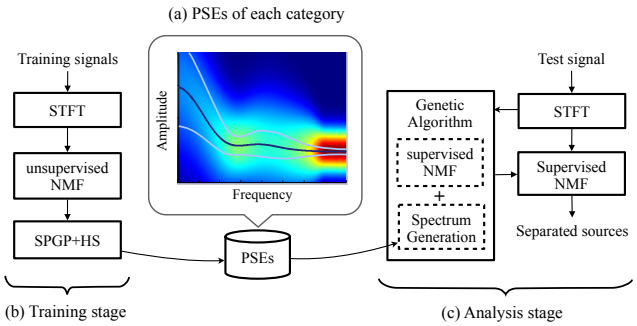

図 1　PSE 法による信号解析のフローチャート．(a) は確率スペクトル包絡 (PSE) を表す．赤色，青色はそれぞれ高い確率値，低い確率値を示している．黒線は平均曲線，その上下の白線は平均曲線 ± 分散曲線をプロットしたものである．

Fig. 1　Flowchart of signal analysis using PSEs. (a) Propabilistic spectrum envelope (PSE): the red and blue color indicate the large and small values of probability, respectively. The black line is a mean envelope, and white lines are mean plus/minus variance envelopes.

従来の PSE に基づく手法 [12] では，遺伝アルゴリズムを用いて最適なスペクトル集合を探索している．しかしながら，PSE はスペクトル生成の自由度が高く探索空間が広いため，意図しないスペクトルまで生成が可能となってしまう．そのため，オクターブの違いや，カテゴリの違いを区別することが困難であった．そこで本研究では，オクターブの違い，カテゴリの違いを区別するため，アクティビティ行列のスパース性，凝集性に関する制約項を付与した新たな評価関数を定義し，より分離精度の高い信号解析を目指す．

2. 提案手法の概要

この章では，我々がこれまでに提案してきた確率スペクトル包絡 (probabilistic spectrum envelope; PSE) を用いた信号解析手法の概要について述べる．PSE 法では，楽器や音素など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，教師あり NMF を繰り返すことで，未知のイベントに対応した音源分離を実現する．提案手法のフローチャートを図 1 に示す．

PSE 法は，確率スペクトル包絡を求める学習ステップと，実際に音源分離を行う解析ステップに分かれる．学習ステップでは，まず学習データのスペクトログラムを計算する．ここで学習データは，カテゴリごとに用意され，各イベントがそれぞれ独立に演奏された（すなわち同時には演奏されない）音響信号を用いる．また本研究では，有声音や楽器音など，基本周波数が存在するイベントを対象としており，無声音やドラム音など，

基本周波数が存在しないものは取り扱わない．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イベント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拡張した SPGP+HS [13] により，スペクトル包絡の平均と分散を学習する．この平均と分散によって表現される包絡線を，確率スペクトル包絡 (PSE) と呼ぶ．以上の手順により，カテゴリ毎に PSE を求める．

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) から計算される評価関数を最小化するように，遺伝アルゴリズムを用いて，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，最終的な音源分離の結果とする．

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，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．

- 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，

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

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

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 [14]．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 [15] 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 [13] 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．

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

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

ただし，$\mathbf{Q} = \mathbf{K}_{f_mf'_m} + \mathbf{K}_{f_mf_n}(\boldsymbol{\Lambda}+\sigma_\lambda^2\mathbf{I})^{-1}\mathbf{K}_{f_nf_m} + diag(\boldsymbol{h})$, $\acute{\mathbf{Q}} = (\mathbf{K}_{f_mf'_m} + diag(\boldsymbol{h}))^{-1} - \mathbf{Q}^{-1}$，$\boldsymbol{\Lambda} = diag(\mathbf{K}_{f_nf_n} - \mathbf{K}_{f_nf_m}\mathbf{K}_m^{-1}\mathbf{K}_{f_mf_n})$ 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 [13] を参照されたい．

学習ステップでは，以上の手順をカテゴリ毎に行う．式 (5),(6) では $\mu(f)$ と $\sigma(f)$ を一般化して表記しているが，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 はカテゴリの数である．

4. 解析ステップ

4.1 PSE に基づくスペクトルのランダム生成

カテゴリ c の確率スペクトル包絡 (PSE) に基づくスペクトル包絡 $e^c(f)$ は，次式のようにランダムに生成される．

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

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 $e^c(f)$ に沿った，基本周波数 f_0 のスペクトル $p(f)$ は

$$p(f) = \max\big(e^c(f),0\big)\cdot\Psi(f;f_0) \tag{8}$$

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．$\Psi(f; f_0)$ は基本周波数 f_0 のくし形調波フィルタであり，式 (9) で計算される．

$$\Psi(f; f_0) = \sum_l \exp\left\{-\frac{(f - f_0 \cdot l)^2}{2\nu^2}\right\} \qquad (9)$$

ここで l はコンポーネントを示すインデックス，ν 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，カテゴリ c，基本周波数 f_0 のスペクトルをランダムに生成することができる．カテゴリ，基本周波数を変えた複数のスペクトルの集合を NMF の基底行列 $\tilde{\mathbf{W}}$ として用いることで，教師あり NMF の枠組みで信号を解析する．

4.2 遺伝アルゴリズムによる最適基底探索

解析ステージで行うべきことは，テストデータに最も適した NMF 行列 $\hat{\mathbf{W}}$ と $\hat{\mathbf{H}}$ を見つけ出すことである．PSE 法では，これらの最適な行列を探索するため遺伝アルゴリズムを用いる．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 (評価値) の高い (小さい) 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．

この解探索法では，まず，U 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 (12) 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 U 個になるまで繰り返す．

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．

ここで，基底行列 $\tilde{\mathbf{W}}_u$ を選択する確率を q_u とすると，

$$q_u = \frac{\frac{1}{\Theta(\tilde{\mathbf{W}}_u)}}{\sum_{u=1}^{U} \frac{1}{\Theta(\tilde{\mathbf{W}}_u)}} \qquad (10)$$

と定義する．ただし，$\Theta(\tilde{\mathbf{W}})$ は $\tilde{\mathbf{W}}$ の評価値 (4.3 節) を表し，$\tilde{\mathbf{W}}$ が適切であるほど $\Theta(\tilde{\mathbf{W}})$ は小さくなるような関数である．したがって，U 個の基底行列の中から，適切な $\tilde{\mathbf{W}}_u$ が選択されやすくなる．

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．

4.3 ランダム基底行列の評価関数

4.3.1 NMF の行列誤差に基づく指標

解析したいテストデータの振幅スペクトログラムを $\mathbf{X}$，ランダムに生成された基底行列 (4.1 節) を $\tilde{\mathbf{W}}$ とすると，次で示されるような擬似逆行列を用いてアクティビティ行列 $\tilde{\mathbf{H}}$ を計算することがことができる．

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．

アクティビティ行列 $\tilde{\mathbf{H}}$ はこのように一意に求まるので，$\tilde{\mathbf{H}}$ は $\tilde{\mathbf{W}}$ の関数とみなすことができる．もし $\tilde{\mathbf{W}}$ のそれぞれの列ベクトルが，与えられたテストデータを構成するスペクトルに近ければ，$\mathbf{X}$ と行列積 $\tilde{\mathbf{H}}\tilde{\mathbf{W}}$ は小さくなるはずである．したがって，行列誤差

$$D_{EUC}(\mathbf{V}, \tilde{\mathbf{W}}\tilde{\mathbf{H}}) = (\mathbf{V} - \tilde{\mathbf{W}}\tilde{\mathbf{H}})^2 \qquad (11)$$

は最適な行列 $\hat{\mathbf{W}}$ と $\hat{\mathbf{H}}$ を探索する指標となる．

4.3.2 制約項を加えた評価関数

従来の PSE 法では，前述の行列誤差をそのまま遺伝アルゴリズムの評価値に用いていた [12]．しかし，本研究ではより適切な基底行列を探索するため，2 つの制約項 (スパース性 $sp(\tilde{\mathbf{H}})$，凝集性 $den(\tilde{\mathbf{H}})$ に関する制約) を追加した新たな評価関数 $\Theta(\tilde{\mathbf{W}})$ を定義する (式 (12))．

$$\Theta(\tilde{\mathbf{W}}) = D_{EUC}(\mathbf{X}, \tilde{\mathbf{W}}\tilde{\mathbf{H}}) - \alpha \cdot sp(\tilde{\mathbf{H}}) - \beta \cdot den(\tilde{\mathbf{H}}) \quad (12)$$

ここで，α ($\geqq 0$) と β ($\geqq 0$) はそれぞれスパース制約，凝集制約の効果を決める重みパラメータであり，本研究では $\alpha = 0.5, \beta = 0.0001$ とした．

スパース制約はアクティビティ行列 $\tilde{\mathbf{H}}$ をスパースに導くための制約項であり，スパース性 $sp(\tilde{\mathbf{H}})$ は次のように計算される．

$$sp(\tilde{\mathbf{H}}) = \frac{\#\left\{(j,k)|\tilde{\mathbf{H}}_{jk} \leqq \epsilon\right\}}{R \times T} \qquad (13)$$

ここで，ϵ ($\geqq 0$) は微小値をとる定数である (実験では，$\epsilon = 0.1$ とした)．図 2 の (a), (b), (c) はそれぞれ PSE によって生成された，基本周波数が $f_0, f_0, 2f_0$ のスペクトルを模したものであり，(b) と (c) を加算したものは (a) に等しい．しかしながら，基底ベクトルが (a) の場合，(b),(c) の場合とでは式 (11) で計算される行列誤差が等しくなる．そのため，このような指標ではオクターブの違いを区別することができない．そこで，提案法では式 (13) のスパース制約を加えることで，倍音を基本周波数とする基底ベクトルの出現を抑える．

また，PSE 法にはもう一つの問題点がある．図 3 に示すように，異なる PSE から全く同じスペクトルを生成し得るので，このような事例ではカテゴリの区別ができない．この問題を解

決するため，本稿ではスパース性に加えて凝集性 $den(\tilde{\mathbf{H}})$ の制約項を提示する．式 (14) で定義される凝集性は，アクティビティ行列のアクティブ要素が時間，または行インデックス（すなわち基本周波数とカテゴリ）に関して近くに存在すればするほど大きな値となる．

$$den(\tilde{\mathbf{H}}) = \frac{\sum_{k,l,l'} \exp\left\{-\frac{(s_{k,l}-s_{k+1,l'})^2}{2\rho^2}\right\}}{\sum_k N_k} \tag{14}$$

$$\{s_{k,l}\}_{l=1}^{N_k} = \left\{ j \mid \tilde{\mathbf{H}}_{jk} \geq \epsilon \right\} \tag{15}$$

ここで ρ は行インデックスの距離範囲を決める定数 (我々の実験では $\rho = 3$) である．$s_{k,l}$ は列ベクトル $\tilde{\mathbf{H}}_{:k}$ の l 番目のアクティブ要素 (微小値以上の値を持つ行列要素)，N_k は列ベクトル $\tilde{\mathbf{H}}_{:k}$ のアクティブ要素の数を表す．この凝集制約を用いることで，アクティブ要素が空間的に近くに集まりやすくなり，アクティビティ行列の断片化，誤推定を防ぐことができる．

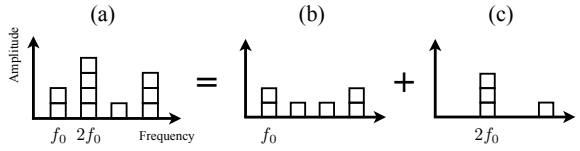

図 2　スパース制約導入の動機．

Fig. 2　Motivation for introducing a sparseness constraint.

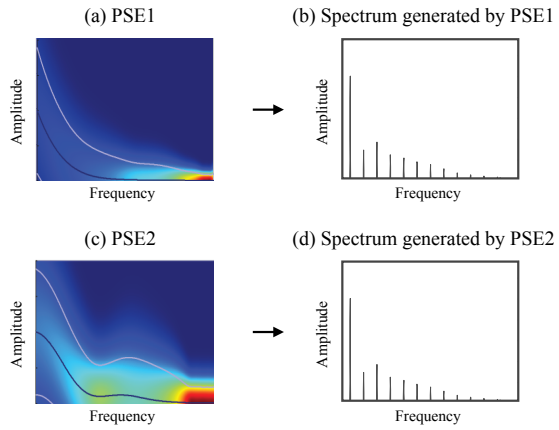

図 3　凝集制約導入の動機

Fig. 3　Motivation for introducing a density constraint.

4.4　事後処理

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 $\hat{\mathbf{H}}$ を最終的な解析結果とする．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 $\hat{\mathbf{W}}$ には楽器や音素など，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，発音長，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　実験条件

提案手法の有効性を確かめるため，音楽音響信号を対象とした評価実験を行った．この実験では，MIDI 音源で演奏された音響信号を提案法によって自動採譜し，再び MIDI 形式へ変換したときの変換精度を測る．提案手法では予め楽器カテゴリごとに確率スペクトル包絡 (PSE) の学習を行う．今回の実験ではピアノとフルートの 2 種類の楽器カテゴリの PSE を学習させた．このときに用いた MIDI 音源はそれぞれ 48 音階分 (C2∼B5) の “Piano1”, 24 音階分 (C4∼B5) の “Flute1” を用いた．解析対象となるテストデータとして，RWC データベース(注1)の “RWC-MDB-C-2001 No. 43: Sicilienne op.78”の一部 (図 5 (a)) を．2 種の楽器 (ピアノとフルート) で演奏させ，モノラルで録音させたものを用いている．ここで，ピアノとフルートの音源にはそれぞれ MIDI 音源の “Piano2”，“Flute2” を用いている．これらの音源は，PSE を学習させるときに用いた音源とは異なる．

提案法によって得られた最適なアクティビティ行列 $\hat{\mathbf{H}}$ を適切な閾値によって 2 値化し，最終的に MIDI フォーマットへ変換した．このとき，遺伝アルゴリズムの評価関数にスパース性，凝集性に関する制約項を含めた場合，含めない場合で結果を比較した (それぞれ “sp+den”, “sp”, “den”, “w/o”)．これらは解の探索に遺伝アルゴリズムによるランダム探索を用いているため，結果は初期値 $\{\tilde{\mathbf{W}}_u\}_{u=1}^{U}$ や突然変異行列に依存する．そのため初期値を固定し，それぞれの手法で 100 回解析を繰り返し，正解率の平均，最大値，最小値を算出した．また，PSE を用いない従来の信号解析手法である教師あり NMF の結果とも比較を行った．解析時に正解である “Piano2”, “Flute2” を基底行列に与えた場合を “ideal”，PSE 法と条件が等しい “Piano1”, “Flute1” を与えた場合を “s-NMF” とする．

5.2　実験結果と考察

図 4 に，各手法による音符正解率を示す．音符正解率は，

$$\frac{N_{all} - (N_{ins} + N_{del})}{N_{all}} \times 100 \tag{16}$$

で計算される．ただし，N_{all}，N_{ins}，N_{del} はそれぞれ音符の総数，音符の挿入誤り数，音符の削除誤り数を表す．前述の 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 $\tau = 0.3$ [sec.] とした．図 4 の PSE 法における棒グラフの値は 100 回の解析の平均，エラーバーは最大値，最小値を表す．教師あり NMF の結果を見ると，システムがテストデータに用いられている音源を知っている場合 (“ideal”)，高い正解率を示している．しかしながら，与える基底行列がテストデータの音源とは異なる場合 (“s-NMF”) 音符正解率は極端に落ちている．一方，PSE 法による結果 (“sp+den”, “sp”, “den”, “w/o”) では，システムが

(注1)：http://staff.aist.go.jp/m.goto/RWC-MDB/

その音源を知らないにも関わらず，正解率はあまり低下しない．この望ましい結果は PSE 法によって未知の楽器機種のスペクトルを生成できたためであると考えられる．PSE 法の結果の中で比較すると，評価関数に制約項を加えない場合よりも，スパース制約，凝集制約，またはそれら両方を加えた方が高い正解率が得られた．中でも両方の制約を加えた場合 (“sp+den”) が最も良い結果を示しており，100 回の解析の中では理想値である “ideal” をも上回る試行が存在した．これは，アクティビティ行列をスパースに導くスパース制約と，アクティブ要素が時間-音階空間の中で散布しないようにする凝集制約の相乗効果によって正解率が向上したと考えられる．

図 5 (b) は手法 “sp+den” による解析結果の一例を示している．ほとんどの音符は正しく推定されているが，一部でオクターブ違いの誤りが生じている．今後更なる制約項を加えることでこうした誤りを軽減していきたい．

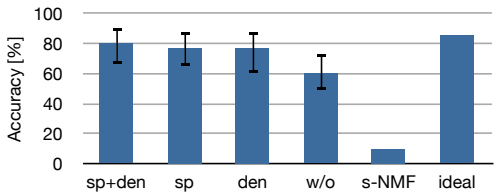

図 4　各手法による音符正解率．
Fig. 4　Accuracy rates of each method.

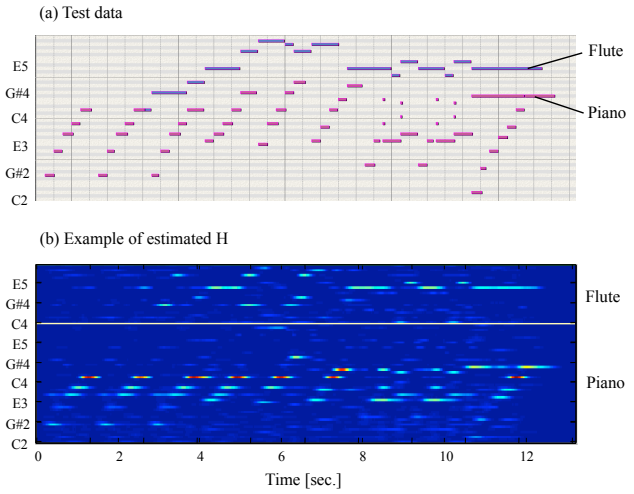

図 5　(a) テストデータのピアノロール．赤と紫の音符はピアノ音，ヴァイオリン音を表す．(b) スパース制約，凝集制約を加えた場合の解析結果例．
Fig. 5　(a) Piano-roll representation of test MIDI data. The red and purple parts indicate piano and violin tones, respectively. (b) Example of analysis results with sparseness and density constraints.

6. おわりに

本論文では，モノラル信号の信号解析・音源分離手法に関するアルゴリズムを提案した．従来の PSE 法では NMF 行列の積と観測スペクトログラムの差を最小化するように最適な NMF 行列を探索していたが，PSE によるスペクトル生成の自由度が高いため，最適な解を得ることは困難であった．本稿では評価関数にスパース性，凝集性に基づく制約項を加えることにより，最適な解の探索を促す方法を提案した．評価実験により，提案法による解析結果が最も最適解に近いことを実証した．

文　献

[1] O.L. Frost, “An algorithm for linearly constrained adaptive array processing,” Proceedings of the IEEE, vol.60, pp.926–935, 1972.

[2] Y. Mori, H. Saruwatari, T. Takatani, S. Ukai, K. Shikano, T. Hiekata, Y. Ikeda, H. Hashimoto, and T. Morita, “Blind separation of acoustic signals combining simo-model-based independent component analysis and binary masking,” EURASIP J. Appl. Signal Process., vol.2006, pp.194–194, Jan. 2006.

[3] S.T. Roweis, “One microphone source separation,” In Advances in Neural Information Processing Systems 13, pp.793–799, MIT Press, 2000.

[4] G. jinJang and T. wonLee, “A maximum likelihood approach to single-channel source separation,” Journal of Machine Learning Research, vol.4, pp.1365–1392, 2003.

[5] P. Smaragdis and J.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.

[6] T. Virtanen, “Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria,” Audio, Speech, and Language Processing, IEEE Transactions on, vol.15, no.3, pp.1066–1074, 2007.

[7] O. Dikmen and A.T. Cemgil, “Unsupervised single-channel source separation using bayesian nmf,” Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA’09. IEEE Workshop on, pp.93–96, 2009.

[8] M.N. Schmidt and R.K. Olsson, “Single-channel speech separation using sparse non-negative matrix factorization,” International Conference on Spoken Language Processing (INTERSPEECH), vol.2, p.1, 2006.

[9] A. Cont, S. Dubnov, and D. Wessel, “Realtime multiple-pitch and multiple-instrument recognition for music signals using sparse non-negative constraints,” Proceedings of Digital Audio Effects Conference (DAFx), pp.10–12, 2007.

[10] J.F. Gemmeke and T. Virtanen, “Noise robust exemplar-based connected digit recognition,” Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, pp.4546–4549, 2010.

[11] 中鹿　亘，滝口哲也，有木康雄，“基底の反復生成と教師あり nmf を用いた信号解析，” 電子情報通信学会技術研究報告，vol.110, no.357，pp.195–200，2010.

[12] 中鹿　亘，滝口哲也，有木康雄，“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，” 情報処理学会研究報告，vol.2011，no.18，pp.1–6，2011.

[13] E. Snelson and Z. Ghahramani, “Variable noise and dimensionality reduction for sparse Gaussian processes,” Proceedings of the 22nd International Conference on Uncertainty in Artificial Intelligence, pp.1–8, 2006.

[14] D.D. Lee and H.S. Seung, “Algorithms for non-negative matrix factorization,” Advances in neural information processing systems, vol.13, pp.1–7, 2001.

[15] X. Huang, A. Acero, and H.W. Hon, Spoken language processing: A guide to theory, algorithm, and system development, Prentice Hall PTR Upper Saddle River, NJ, USA, 2001.