Vol.2011-MUS-91 No.5
2011/7/27

時間周波数分解能の異なるスペクトログラムの並列NMFによる多重音解析

落合和樹[†1]　中野允裕[†1]
小野順貴[†1]　嵯峨山茂樹[†1]

本報告では，自動採譜のためにスペクトログラムの非負値行列分解 (NMF) を用いた新しい多重音解析手法を提案する．自動採譜には音高の推定と発音時刻の推定が同時に必要であるが，音高と発音時刻の推定には解析フレーム長に関するトレードオフが存在する．そこで，異なる解析フレームによるスペクトログラムを併用することにより，双方の分解能を保ち音高と発音時刻の推定精度を高めることができると考えられることから，高時間分解能と高周波数分解能の 2 種類のスペクトログラムに対して NMF を並列的に用いる方法を提案し，実際の音響信号に対し簡単な発音検出実験を行い，従来の NMF と比較をしその有用性を示す．

Concurrent nonnegative matrix factorization using multi-resolution spectrograms for multipitch analysis of music signals

Kazuki Ochiai,[†1] Masahiro Nakano,[†1]
Nobutaka Ono[†1] and Shigeki Sagayama[†1]

The Short-Time Fourier Transform (STFT) is commonly used as a back-end to multipitch analysis and it transforms acoustic signals into the time-frequency representation. However, a trade-off between time and frequency resolutions exists, depending on analysis frame length. It is therefore difficult to simultaneously obtain high accuracy of both note onset and note pitch estimation. Combination of spectrograms obtained with different frame lengths can achieve high resolution on both time and frequency, which should improve note parameter estimation. In this paper we propose a new method of multipitch analysis of musical signals based on Nonnegative Matrix Factorization (NMF) of spectrograms, where the NMF is applied in parallel to high time resolution and high frequency resolution spectrograms. We demonstrate the efficiency of our approach through note detection experiments.

1. はじめに

自動採譜は音楽音響信号処理における重要な課題の 1 つであり，即興演奏などの録音データしかない曲を楽譜にすることで練習に役立てたり，音楽を音符の記号列に変換することで MIDI 変換や音楽検索などに応用したりできる．自動採譜のためにはどの高さの音がどのタイミングで鳴り始めどれくらいの時間長で鳴っているかという音符情報の取得と，得られた音符列を音楽的に正しい楽譜に作り上げることが必要である．本研究ではこれらのうち単一楽器で演奏された多重音中の音符検出に着目する．

多重音から基本周波数や発音 (オンセット) 時刻を推定する研究は従来から数多くなされており，近年では，多重音解析に有効な手段として非負値行列分解 (Nonnegative Matrix Factorization; NMF) が注目されている[1)]．これは，ある単音を 1 つの基底スペクトルでモデル化しそれが音量のみ変化しているとみなし，スペクトログラムがスパースであるという仮定に基づいて多重音を単音毎に分解できることを期待している．NMF による解析性能向上のため，各音の時間連続性[2)] や楽器音の調波構造性を利用した手法[3)] など，単一のスペクトログラムを西洋音楽的に解釈した制約が考えられてきた．しかし，性能は飽和気味であり，更なる性能向上のためには新たな視点が必要である．

時間分解能と周波数分解能の間には不確定性原理に基づく解析フレーム長に関するトレードオフが存在している．一般に信号はある時間長のフレームに区切られて解析される．このとき，音高の推定精度を向上させるには高周波数分解能を実現できる長い解析フレームが必要となる．例えば，A0 (27.5 Hz) と A♯0 (29.1 Hz) の区別には約 1.5 Hz の周波数分解能が求められる．一方，発音時刻の推定精度を高めるには高時間分解能を実現できる短い解析フレームが必要となる．例えば，テンポ 200 bpm で連続する 16 分音符の区別には約 60 ms の時間分解能が求められる．このため，音高と発音時刻の高精度な推定を同時に実現できる 1 つの最適なフレーム長を求めることは難しい．そこで，異なるフレーム長で解析された複数のスペクトログラムを併用することで時間周波数分解能を同時に高めた解析ができると考えられる．これに関しては 2 つのスペクトログラムから確率潜在コンポーネント解析 (Probabilistic Latent Component Analysis; PLCA) を用いて時間周波数分解能の不確

†1 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

Vol.2011-MUS-91 No.5
2011/7/27

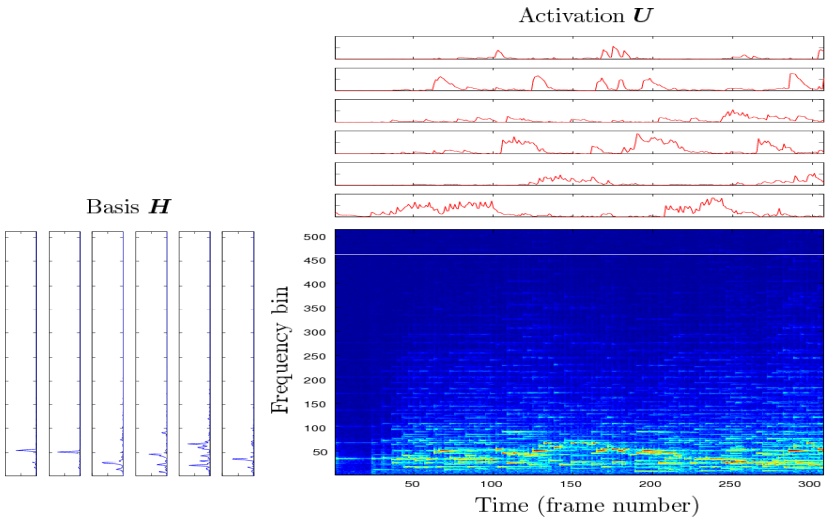
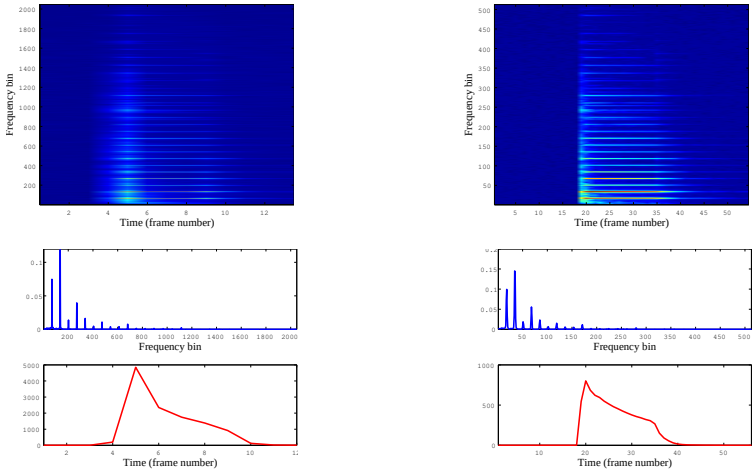


図 1 音楽音響信号のスペクトログラムに NMF を適用した例．基底スペクトル行列 $\boldsymbol{H}$ とそのアクティベーション行列 $\boldsymbol{U}$ に分解される．
Fig. 1 An example of the regular NMF applied to a musical signal. The spectrogram is decomposed into spectral basis and its activation matrices.



(a) フレーム長：256 ms　(b) フレーム長：64 ms

図 2 異なるフレーム長で解析されたピアノの C4 音のスペクトログラム (上) と，それぞれ NMF により推定された基底スペクトル (中) とそのアクティベーション (下)．解析フレームが長いとスペクトルの形状が鋭くなるが，オンセットとオフセットのタイミングが曖昧になる．一方，解析フレームが短いとアクティベーションの形状ははっきりするが，スペクトルはぼやける．お互いに等価なスペクトルとアクティベーションが得られることが期待される．
Fig. 2 Spectrograms of a single note signal (C_4) analyzed with different frame lengths (top) and the estimated spectral basis and activation matrices (middle and bottom). When the long frame is used, the spectral shape is sharp while the note onset and offset timing are ambiguous. On the other hand, the short frame provides clear activation change and blurred spectrum. It is expected to obtain equivalent basis and activation matrices from them.

定性原理を超えたスペクトログラムを生成する手法が提案されている[4)]．本報告では同様にこのトレードオフを解消するため，2 つの異なるフレーム長でのスペクトログラムを並列に NMF で分解する新しい音楽音響信号分解手法を提案し，実際に発音検出実験を行い有用性を検討する．

2. 提 案 手 法

2.1 スペクトログラムの並列 NMF

本報告では短時間 Fourier 変換 (STFT) によって得られたスペクトログラムを扱う．NMF によるスペクトログラムの分解表現は，観測された振幅 (もしくはパワー) スペクトログラムを非負値行列 $\boldsymbol{Y} = (Y_{\omega,t}) \in \mathbb{R}^{\geq 0, \Omega \times T}$ とみなし，これが限られた数の基底の重ね合わせで表現されるという仮定の下，

$$Y_{\omega,t} \simeq \sum_{i=1}^{I} H_{\omega,i} U_{i,t} \tag{1}$$

となるような基底行列 $\boldsymbol{H} = (H_{\omega,i}) \in \mathbb{R}^{\geq 0, \Omega \times I}$ とアクティベーション行列 $\boldsymbol{U} = (U_{i,t}) \in \mathbb{R}^{\geq 0, I \times T}$ を決定することで得られる (図 1)．ここで，$\omega = 1, \cdots, \Omega$ は周波数ビンのインデックス，$t = 1, \cdots, T$ は時刻に対応するインデックス，$i = 1, \cdots, I$ は基底のインデックスであり，観測スペクトログラムが I 個の基底スペクトルと各基底の音量に相当するアクティベーションの積で表現されるというモデルとなっている．

提案手法では音高と発音時刻の推定精度を両立するために，異なるフレーム長で解析された 2 つのスペクトログラム $\boldsymbol{Y}^{(S)}, \boldsymbol{Y}^{(L)}$ を併用して STFT における不確定性原理による時間周波数分解能のトレードオフを解消すること，また，その上でそれぞれの基底スペクト

ルとアクティベーションのペアが楽器音の各 1 音高に対応することを狙っている．ここで，添え字の S は短いフレーム，L は長いフレームで解析されたものを表す．それぞれのスペクトログラムに対し独立に NMF を適用し，得られた基底とアクティベーションを組み合わせて音符検出をすることも考えられるが，それぞれ解析された基底とアクティベーションの対応関係が取れずにうまくいかない場合がある．例えば，短いフレームでのスペクトログラムを用いるときに，周波数分解能が低いために 1 つの周波数ビンに 2 つの音高の基本周波数が入ってしまうことがある．この場合に NMF を行うと，基底とアクティベーションの反復推定 (次節参照) の際に時間分解能が高いアクティベーション側にも誤推定を生んでしまう．これに対して周波数分解能の高いスペクトログラムから得られる基底の情報を参照しながら更新することでこの問題を回避できると考えられる．また，1 つの信号を異なる条件下で解析しているだけなので，NMF で得られる基底とアクティベーションは同一であるべきである (図 2)．

そこで，NMF のパラメータ推定にそれぞれのスペクトログラムから得られる基底とアクティベーションの形状が類似しているという正則化を加えることでこういった誤推定を抑制できる．本報告では形状類似性の正則化項を次式のような対応フレーム，周波数ビン間の二乗誤差とし，

$$\mathcal{R}_H(\theta) = \sum_{i}^{I} \sum_{\omega_S}^{\Omega_S} \left| H^{(S)}_{\omega_S,i} - \sum_{\omega_L \in \omega_S} H^{(L)}_{\omega_L,i} \right|^2 \tag{2}$$

$$\mathcal{R}_U(\theta) = \sum_{i}^{I} \sum_{t_L}^{T_L} \left| U^{(L)}_{i,t_L} - \sum_{t_S \in t_L} U^{(S)}_{i,t_S} \right|^2 \tag{3}$$

のように定義する．ここで，$\theta = \{\boldsymbol{H}^{(S)}, \boldsymbol{H}^{(L)}, \boldsymbol{U}^{(S)}, \boldsymbol{U}^{(L)}\}$ である．

それぞれの基底スペクトルとアクティベーションのペアを楽器音の各 1 音高に対応させる点に関しては，楽器音には基本周波数と倍音に強いエネルギーを持つという性質を利用する．NMF では，単音毎に調波成分のみ非零とし乗法更新によりその構造を保持する手法[3] や，調波成分を複数の調波構造の線形和で表現する手法[5] が提案されている．この他にも，基本周波数 ω_i とその倍音 $k\omega_i (k = 2, \cdots, K)$ に小さい分散 σ を持つ正規分布の重み $a_{i,k}$ での混合で単音のスペクトルを表すというモデルが提案されている[6]．本研究においてもこれらの枠組みは利用できると考えられ，単音に分離するために次式のような打ち切りの正規分布の混合を基底の初期値とする．

$$H^{(n)}_{\omega_n,i} = \begin{cases} \displaystyle\sum_{k=1}^{K} \frac{a_{i,k}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\omega_n - \log k\omega_i)^2}{2\sigma^2} \right] & \left(2^{-\frac{1}{12}} k\omega_i \le \omega_n < 2^{\frac{1}{12}} k\omega_i \right) \\ 0 & (\text{otherwise}) \end{cases} \tag{4}$$

ここで，以下添え字の n は $n = \{S, L\}$ を表すこととする．

また，分解スケールの任意性を回避するため，

$$\sum_{\omega_n} H^{(n)}_{\omega_n,i} = 1 \qquad (i = 1, \cdots I) \tag{5}$$

を仮定する．

2.2 最適化アルゴリズム

NMF は一般的に観測とモデル間の何らかの距離尺度を目的関数とし，これを最小化する制約付き最適化問題として解かれる．目的関数を解析的に最適化することは困難であり，主に反復計算によりパラメータを更新する方法が用いられる．距離尺度としては二乗誤差や I ダイバージェンス，板倉斎藤距離などがよく用いられており，いずれにおいても，効率のよい乗法更新アルゴリズムにより非負性の保証された解が得られることがわかっている[1),7)]．

本報告では距離尺度を I ダイバージェンス

$$\mathcal{I}(\theta) = \sum_{\omega,t} \left[Y_{\omega,t} \log \frac{Y_{\omega,t}}{\sum_i H_{\omega,i} U_{i,t}} - \left(Y_{\omega,t} - \sum_i H_{\omega,i} U_{i,t} \right) \right] \tag{6}$$

とした場合における最適化アルゴリズムを考える．このとき，解くべき問題は観測されたスペクトログラム $\boldsymbol{Y}$ から

$$\begin{aligned} &\text{minimize} \quad \mathcal{J}(\theta) = \sum_n \mathcal{I}^{(n)}(\theta) + \mu_H \mathcal{R}_H(\theta) + \mu_U \mathcal{R}_U(\theta) + \lambda \mathcal{S}(\theta) + \eta \mathcal{Q}(\theta) \\ &\text{subjectto} \quad \forall_i \sum_{\omega_n} H^{(n)}_{\omega_n,i} = 1, \quad \forall_{\omega_n,i} H^{(n)}_{\omega_n,i} \ge 0, \quad \forall_{i,t_n} U^{(n)}_{i,t_n} \ge 0, \\ &\qquad n = \{S, L\}, \quad \mu_H,\ \mu_U,\ \lambda,\ \eta \ge 0 \end{aligned} \tag{7}$$

を与える θ を求める問題となる．ここで，$\mathcal{S}(\theta)$ はスパースな解へ誘導する正則化項であり，アクティベーションに関して L_p ノルム

$$\mathcal{S}(\theta) = \sum_{i,t_n,n} \left| U^{(n)}_{i,t_n} \right|^p \qquad (0 < p \le 1) \tag{8}$$

とする[8)]．$\mathcal{Q}(\theta)$ は半音異なる基底が表すスペクトルの倍音構造が類似しているという仮定に関する正則化項であり，

$$\mathcal{Q}(\theta) = \sum_n \left\| \boldsymbol{H}^{(n)} - \boldsymbol{W}^{(n)} \boldsymbol{H}^{(n)} \boldsymbol{V}^{(n)} \right\|_2^2 \tag{9}$$

とする[3]．$\boldsymbol{W}^{(n)}$ は各基底スペクトルを半音分上げる変換行列で，$\boldsymbol{V}^{(n)}$ は各基底を 1 列右にシフトさせる行列である．また，μ_H, μ_U, λ, η はそれぞれの正則化項に関する定係数である．

この問題を解くアルゴリズム導出のために補助関数法を用いる[9]．I ダイバージェンス $\mathcal{I}(\theta)$ 及びスパース正則化項 $\mathcal{S}(\theta)$ に関する補助関数 $\mathcal{I}^{+(n)}(\theta, \boldsymbol{\xi}^{(n)}), \mathcal{S}^{+}(\theta, \boldsymbol{U}'^{(S)}, \boldsymbol{U}'^{(L)})$ はそれぞれ

$$\mathcal{I}^{(n)}(\theta) \leq \mathcal{I}^{+(n)}(\theta, \xi^{(n)}) = \sum_{\omega, t_n} \left[Y^{(n)}_{\omega_n, t_n} \log Y^{(n)}_{\omega_n, t_n} - Y^{(n)}_{\omega_n, t_n} + \sum_i H^{(n)}_{\omega_n, i} U^{(n)}_{i, t_n} - Y^{(n)}_{\omega_n, t_n} \sum_i \xi^{(n)}_{\omega_n, t_n, i} \log \frac{H^{(n)}_{\omega_n, i} U^{(n)}_{i, t_n}}{\xi^{(n)}_{\omega_n, t_n, i}} \right] \tag{10}$$

$$\mathcal{S}(\theta) \leq \mathcal{S}^{+}(\theta, \boldsymbol{U}'^{(S)}, \boldsymbol{U}'^{(L)}) = \sum_{i, t_n, n} p \left| U'^{(n)}_{i, t_n} \right|^{p-1} \left(U^{(n)}_{i, t_n} - U'^{(n)}_{i, t_n} \right) + \left| U'^{(n)}_{i, t_n} \right|^{p} \tag{11}$$

と設計できる．ここで，$\boldsymbol{\xi}^{(n)}$ と $\boldsymbol{U}'^{(n)}$ は補助変数 $\hat{\theta}\,(= \{\xi^{(S)}_{\omega_S, t_S, i}, \xi^{(L)}_{\omega_L, t_L, i}, U'^{(S)}_{i, t_S}, U'^{(L)}_{i, t_L}\})$ であり，$\boldsymbol{\xi}^{(n)}$ は

$$0 < \xi^{(n)}_{\omega_n, t_n, i} < 1, \qquad \sum_i \xi^{(n)}_{\omega_n, t_n, i} = 1 \tag{12}$$

を満たし，$\boldsymbol{U}'$ は 1 ステップ前の更新値とする．式 (10) および (11) の等号は

$$\xi^{(n)}_{\omega_n, t_n, i} = \frac{H^{(n)}_{\omega_n, i} U^{(n)}_{i, t_n}}{\sum_{i'} H^{(n)}_{\omega_n, i'} U^{(n)}_{i', t_n}} \tag{13}$$

$$U'^{(n)}_{i, t_n} = U^{(n)}_{i, t_n} \tag{14}$$

のときに成立する．このとき，式 (7) に式 (10) と (11) を適用した補助関数 $\mathcal{J}^{+}(\theta, \hat{\theta})$ を最小化する θ を求めればよい．そのための更新式は

$$\frac{\partial \mathcal{J}^{+}(\theta, \hat{\theta})}{U^{(S)}_{i, t_n}} = 0, \qquad \frac{\partial \mathcal{J}^{+}(\theta, \hat{\theta})}{U^{(L)}_{i, t_n}} = 0$$
$$\frac{\partial \mathcal{J}^{+}(\theta, \hat{\theta})}{H^{(S)}_{\omega_n, i}} = 0, \qquad \frac{\partial \mathcal{J}^{+}(\theta, \hat{\theta})}{H^{(L)}_{\omega_n, i}} = 0 \tag{15}$$

を θ の各要素について解き，式 (13) と (14) を適用することにより，次式のように得ることができる．

$$U^{(S)}_{i, t_S} \leftarrow U^{(S)}_{i, t_S} \frac{-A_{i, t_S} + \sqrt{A^2_{i, t_S} + 4\mu_U \sum_{t'_S} U^{(S)}_{i, t'_S} \sum_{\omega_S} \frac{Y^{(S)}_{\omega_S, t_S} H^{(S)}_{\omega_S, i}}{X^{(S)}_{\omega_S, t_S}}}}{2\mu_U \sum_{t'_S} U^{(S)}_{i, t'_S}} \tag{16}$$

$$U^{(L)}_{i, t_L} \leftarrow \frac{-B_{i, t_L} + \sqrt{B^2_{i, t_L} + 4\mu_U U^{(L)}_{i, t_L} \sum_{\omega_L} \frac{Y^{(L)}_{\omega_L, t_L} H^{(L)}_{\omega_L, i}}{X^{(L)}_{\omega_L, t_L}}}}{2\mu_U} \tag{17}$$

$$H^{(S)}_{\omega_S, i} \leftarrow \frac{-C_{\omega_S, i} + \sqrt{C^2_{\omega_S, i} + 4D_{\omega_S, i} H^{(S)}_{\omega_S, i} \sum_{t_S} \frac{Y^{(S)}_{\omega_S, t_S} U^{(S)}_{i, t_S}}{X^{(S)}_{\omega_S, t_S}}}}{2D_{\omega_S, i}} \tag{18}$$

$$H^{(L)}_{\omega_L, i} \leftarrow H^{(L)}_{\omega_L, i} \frac{-E_{\omega_L, i} + \sqrt{E^2_{\omega_L, i} + 4F_{\omega_L, i} \sum_{t_L} \frac{Y^{(L)}_{\omega_L, t_L} U^{(S)}_{i, t_L}}{X^{(L)}_{\omega_L, t_L}}}}{2F_{\omega_L, i}} \tag{19}$$

ここで，

$$X^{(n)}_{\omega_n, t_n} = \sum_i H^{(n)}_{\omega_n, i} U^{(n)}_{i, t_n} \tag{20}$$

$$A_{i, t_S} = 1 - \mu_U U^{(L)}_{i, t_L} + \lambda p \left| U^{(S)}_{i, t_S} \right|^{p-1} \tag{21}$$

$$B_{i, t_L} = 1 - \mu_U \sum_{t_S} U^{(S)}_{i, t_S} + \lambda p \left| U^{(L)}_{i, t_L} \right|^{p-1} \tag{22}$$

$$C_{\omega_S, i} = \sum_{t_S} U^{(S)}_{i, t_S} - \mu_H \sum_{\omega_L} H^{(L)}_{\omega_L, i} + G^{(S)}_{\omega_S, i} \tag{23}$$

$$D_{\omega_S, i} = \mu_H + \eta \left[1 + \sum_{\omega'_S} \left(W^{(S)}_{\omega'_S, \omega_S} \right)^2 \right] \tag{24}$$

$$E_{\omega_L, i} = \sum_{t_L} U^{(L)}_{i, t_L} - \mu_H H^{(S)}_{\omega_S, i} + G^{(L)}_{\omega_L, i} \tag{25}$$

$$F_{\omega_L, i} = \mu_H \sum_{\omega'_L} H^{(L)}_{\omega'_L, i} + \eta \left[1 + \sum_{\omega'_L} \left(W^{(L)}_{\omega'_L, \omega_L} \right)^2 \right] H^{(L)}_{\omega_L, i} \tag{26}$$

$$G^{(n)}_{\omega_n, i} = \eta \sum_{\omega'_n} W_{\omega'_n, \omega_n} \left(W_{\omega'_n, \omega_n} H^{(n)}_{\omega_n + 1, i} - H^{(n)}_{\omega'_n, i+1} \right) \tag{27}$$

である．これらの更新式は二次方程式の解の形となっているが，NMF における乗法更新アルゴリズムは保たれている．

3. 評価実験

提案法の有効性を検証するために，実演奏の音楽信号に対して発音検出実験を行った．提案する NMF により得られた基底スペクトルとアクティベーションから各音高に対する発音

時刻を推定する方法と結果について以下に述べる．

3.1 音符検出方法

前節で示した更新式をもとに目的関数が収束したときに，基底行列とアクティベーション行列が決定される．その結果を用いて鳴っている各音高とその発音時刻の音符情報を取得する．その際，各音高については式 (4) を初期値とした調波構造基底を半音毎に与え，毎回のパラメータ更新においても半音異なる基底の調波構造が類似している正則化を与えているので，各基底が対応する音高の推定は容易である．一方，発音時刻推定に関しては様々な方法が考えられる．ここで，各音は強弱をつけて演奏されるが，ある一定の音量以上で鳴っているはずである．また，一度鳴り始めた音は，アタックの瞬間に急激に強くなり，その後徐々に弱くなっていき，リリースすると急に小さくなる，という特徴があると考えられる．

そこで，本報告では単純な方法として，2 つのスペクトログラムでの NMF の結果に対し発音消音に関する閾値を用意する．まず，長い解析フレームでのアクティベーションでは閾値未満の値をすべて 0 にする．それに対し，短い解析フレームでのアクティベーションでは，連続する数フレームで閾値を超えている部分のみ残し他の値を 0 にする．そして，前者で閾値を超えた時に後者で対応するフレームでの値がすべて 0 であればその音高は発音されていないとみなす．以上から，0 でないフレームがあれば発音されたとし，その中で最大値を取るフレームを発音時刻とする．時刻のずれは短いフレームでの 2 フレーム分 (本報告では 128 ms) より大きくずれた場合は誤りとした．次節で述べる比較対象とした従来手法における発音時刻推定は，音高は先述の通り各基底から推定できるので，時間分解能が高いひとつのスペクトログラムから得られたアクティベーションに対して提案手法と同じ閾値を用いた．

3.2 発音検出実験

提案法の発音検出における有効性を検証するために，使用された音高が未知の状態で従来の NMF との比較実験を行った．STFT は，フレーム長 64 ms と 256 ms，フレームシフトはフレーム長の半分 (ハーフオーバーラップ)，解析窓は Hanning 窓という条件で行った．用いた楽曲は RWC クラシック音楽データベース[10] よりピアノ曲 5 曲 (RWC-MDB-C-2001 No. 26, 27, 29, 30, 31) のデータ長約 30 s，サンプリング周波数 16 kHz であった．NMF の基底数は楽曲データに登場するすべての音高が含まれるよう 55 とし，正則化項の各係数は $\mu_H = 0.5$, $\mu_U = 2$, $\lambda = 1$, $p = 0.5$, $\eta = 0.5$ とした．反復回数は従来手法提案手法ともに予備実験の結果収束が認められたため 60 回とした．比較対象の従来手法としては，アクティベーションに関するスパース性 (式 (8)) と，半音異なる基底スペクトルの倍音構造の類似性 (式 (9)) の正則化項のみを目的関数に与えた NMF を用い，スペクトログラムは前節で述べた通り，64 ms のものを用いた．NMF の各パラメータの初期値は基底スペクトル行列は式 (4) を用い，アクティベーション行列は 0 から振幅スペクトログラムの各時間周波数ビンにおける最大値までの値をとるランダム値とした．

表 1 解析に用いたピアノ曲と音符検出における F 値 (%)．“Conv.”と “Prop.”はそれぞれ従来手法と提案手法を表す．

Table 1 Piano pieces used for algorithm evaluation and F-measure in note detection (%). “Conv.” and “Prop.” denote the conventional method and the proposed one.

Composer	Title	Notes	Conv.	Prop.
W. A. Mozart	Variations on “Ah Vous Dirai-je Maman”, K. 265/300e	106	93.0	**93.8**
W. A. Mozart	Piano Sonata in A major, K. 331/300i. 1st mvmt.	105	73.0	**86.0**
F. Chopin	Nocturne in E♭ major, Op. 9, No. 2	124	60.5	**82.4**
F. Chopin	Etude in E major, op. 10-3	162	67.2	**84.4**
R. Schumann	“Träumerei” from Suite (Kinderszenen), op. 15	113	73.5	**78.6**

各楽曲について音高と発音時刻を推定し F 値を求めたものを表 1 に，各音高の発音消音時刻をピアノロールで表示したものを図 3 に示す．その際，閾値に関しては全曲における *Recall* が 90 %以上になるように設定した．正解 MIDI と比較をすると，従来手法では各音の発音時を中心に正解音高と半音ずれた音が誤った音として多数検出されていたが，提案手法では倍音成分のいくつかが誤検出された程度となり，発音時刻推定精度を保ちつつ音高推定精度が向上していることが確認できた．

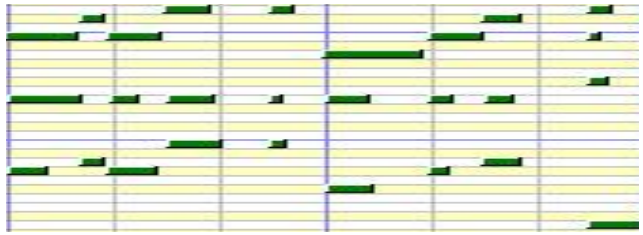
4. おわりに

本報告では，自動採譜に向けた音高と発音時刻の同時推定のために，高時間分解能と高周波数分解能でのスペクトログラムに対して並列に NMF を適用し分解することによる新しい多重音解析手法を提案した．音楽音響信号を用いた採譜実験により，従来手法に比べ音符情報推定精度が向上することを確認した．今回示した結果は限られたデータから得られたものなので，より多くの実験を行う必要がある．今後の課題としては，解析フレーム長の異なる複数のスペクトログラム間にある関係性を捉え，正確な形状類似性に関するモデルを導入することや，リズムモデルを統合することで楽譜を作成するアプリケーションの構築を検討

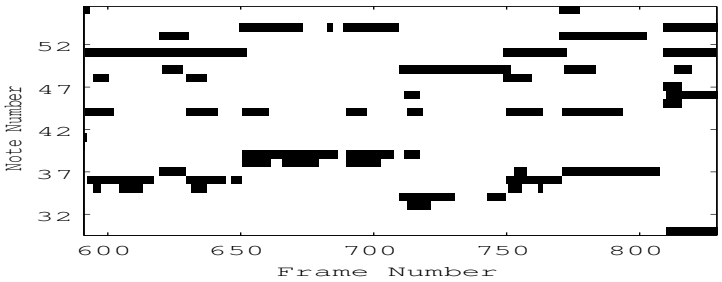
している．

参 考 文 献

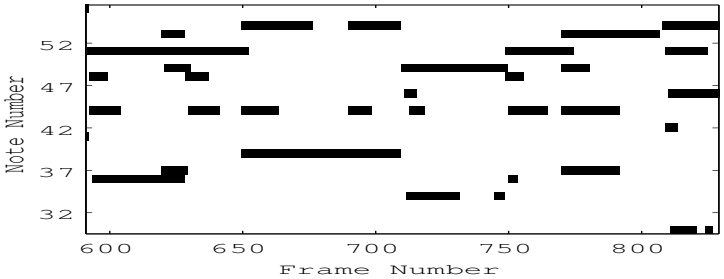
1) Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol.401, pp.788–791 (1999).
2) Virtanen, T.: Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol.15, No.3, pp.1066–1074 (2007).
3) Raczyński, S.A., Ono, N. and Sagayama, S.: Multipitch Analysis with Harmonic Nonnegative Matrix Approximation, *Proc. ISMIR*, pp.381–386 (2007).
4) Nam, J., Mysore, G., Ganseman, J., Lee, K. and Abel, J.S.: A super-resolution spectrogram using coupled PLCA, *Proc. Interspeech*, pp.1696–1699 (2010).
5) Vincent, E., Bertin, N. and Badeau, R.: Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription, *Proc. ICASSP*, pp.109–112 (2008).
6) Kameoka, H., Nishimoto, T. and Sagayama, S.: A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol.15, No.3, pp.982–994 (2007).
7) Févotte, C., Bertin, N. and Durrieu, J.-L.: Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis, *Neural Computation*, Vol.21, No.3, pp.793–830 (2009).
8) Kameoka, H., Ono, N., Kashino, K. and Sagayama, S.: Complex NMF: A new sparse representation for acoustic signals, *Proc. ICASSP*, pp.3437–3440 (2009).
9) Lee, D.D. and Seung, H.S.: Algorithms for Non-negative Matrix Factorization, *Proc. NIPS*, pp.556–562 (2000).
10) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC music database: Popular, classical, and jazz music database, *Proc. ISMIR*, pp.287–288 (2002).



(a) MIDI reference



(b) Conventional NMF



(c) Proposed NMF

図 3 Mozart: Sonata in A Major, K. 331(300i) の正解 MIDI ピアノロールと，提案手法及び従来手法で解析しピアノロールとして表示したもの (一部)．発音されていると推定された音高とその発音から消音までを黒で表示してある．

Fig. 3 MIDI reference and Piano rolls obtained for the conventional and proposed methods applied to the acoustic signal of Mozart's Sonata in A Major, K. 331 (300i).