Vol.2018-MUS-120 No.16
2018/8/22

多重解像度NMFに基づく音響信号演奏詳細解析

保利 武志[1,a)]　中村 和幸[1,b)]　嵯峨山 茂樹[1,c)]

概要：本研究では，音楽音響信号に対する詳細なオンセット時刻や音長，音量を，多重解像度解析に基づき同時に推定する方法について述べる．音楽音響信号からの音響特徴量抽出は自動演奏や自動作曲，音楽情報検索など，近年の学習ベースに基づくモデル獲得の基幹を為すものとして極めて重要な要素技術である．本稿では演奏表情が付与された実演奏ピアノロールは楽譜に基づくピアノロールに対する伸縮により表現可能であるとする仮説のもと，単音ごとにオンセットと音長，音量を陽にモデル化した畳み込み単音重畳モデルに基づく NMF と，さらに複数の異なる時間分解能スペクトログラム解析により得られる基底とオンセット分布，アクティベーション形状を相互に参照し合う並列 NMF へと拡張した多重時間分解能な畳み込み単音重畳 NMF のパラメータ更新式を導出する．提案モデルを用いた音楽音響信号に対するパラメータ推定とその評価実験の結果，畳み込み単音重畳モデルに基づく NMF の有効性が示された．

Multiresolutional NMF for Detailed Audio Analysis of Music Performances

Hori Takeshi[1,a)]　Nakamura Kazuyuki[1,b)]　Sagayama Shigeki[1,c)]

1. はじめに

音楽音響信号からの音響特徴量抽出は自動演奏や自動作曲，音楽情報検索など，近年の学習ベースに基づくモデル獲得の基幹を為すものとして極めて重要な要素技術である．特に人間が楽譜に基づいて演奏する際，多くの場合，楽譜通りの画一的な演奏ではなく，その演奏者の癖や楽譜への解釈などを背景として，アーティキュレーションや緩急，音量変化などが付与された表情豊かなものとなる．

演奏者の演奏モデルを獲得するためには高い時間分解能による詳細な演奏の特徴量解析 (以降，詳細解析と呼ぶ) が必要とされる (例えば楽譜上では和音として表記されている場合でも，実際に演奏する際には同時ではなくわずかにオンセットにずれを生じさせる場合がある)．しかし，フーリエ変換における解析フレーム長に起因する周波数分解能と時間分解能との不確定性原理に基づくトレードオフな関係は，詳細解析におけるボトルネックの要因の一つとなっている．本稿では，表情付き演奏による音楽音響信号に対し，楽譜情報を陽に活用することで精度良く音響特徴量を抽出する方法と，さらに多重解像度分析に基づき周波数分解能及び時間分解能のトレードオフな関係を解消し，オンセット時刻や音量，音長を同時に推定する手法を提案する．

詳細解析において，MIDI 信号からなる演奏データがあれば必要な特徴量の抽出は比較的容易である．しかし，MIDI データを得るためには MIDI ピアノをはじめとした特殊な機材を用いた録音環境が必要であり，また過去の演奏の復元は難しい．加えて，ニューラルネットをはじめとした学習ベースのための特徴量抽出を考えるならば，大量の演奏解析データを必要とされることも多いため，MIDI データだけではなく音楽音響信号に対する解析が広く望まれる．

多重音からなる音楽音響信号の分離問題は，自動採譜問題も含め様々なタスクにおいて手法 [1, 2, 3] が提案され論じられてきたが，近年は非負値行列因子分解 (Non-negative matrix factorization, NMF) [4] を用いた手法が幅広く用いられている [5, 6, 7, 8]．NMF は音響信号スペクトログラムを，頻出パタンからなる基底行列 (音楽音響信号に対し

[1] 明治大学
Meiji University, Nakano, Tokyo 164–8525, Japan
[a)] hori@meiji.ac.jp
[b)] knaka@meiji.ac.jp
[c)] sagayama@meiji.ac.jp

ては，多くの場合各音高の周波数スペクトルに相当) とその時間的音量変化を表すアクティベーション行列との積に分解する．オンセット推定にはアクティベーションに対する閾値処理や隠れマルコフモデル (HMM) などが用いられる．また，前述した周波数分解能と時間分解能のトレードオフを解消する手法として，高周波数分解能なスペクトログラムと高時間分解能なスペクトログラムの 2 つの観測音響信号パワースペクトログラムを用い，基底行列とアクティベーション行列の類似制約を利用した相互に参照し合う並列 NMF [9] も提案されている，

これらの手法の多くは観測音響信号のみから音高列を推定するが，本稿では，実演奏に基づき記述されるピアノロールは，楽譜情報に基づくピアノロールの伸縮により表現されるとする仮説により，楽譜情報を事前知識として利用することを提案する．また，音楽音響信号は各ノート (単音) 個別の音量や音長 (音色) の重畳により表現されるとする単音モデルに基づき，理想的にはオンセット時刻にのみピークがたつオンセット分布と，単音エンベロープ形状パラメータによる畳み込み，そして単音ごとのエネルギーからなるパラメータを含めたモデル化により，オンセット時刻，音長，音量を同時に推定する手法とその多重時間分解能 NMF への拡張モデルを提案する．

2. 畳み込み単音重畳モデルへの定式化

2.1 音楽スペクトログラムのモデル化

ある音楽音響信号が楽譜上に記載可能であるとするならば，その音楽はそれぞれ音量や音長，音色などを属性として持つ単音の組み合わせ (あるいは重畳) により表現可能である．採譜問題を考える場合，音長に関しては例えば四分音符や八分音符など特定の量子化されたもののみを考えれば良いが，表情付き演奏に対する詳細解析を考える場合は，それぞれの音長や音量等を個別に捉える必要がある．

まず，各音高を表すスペクトルパタンを，それぞれ独立した基底ベクトルからなる行列 $\boldsymbol{H}$ で表現することを考える．なお問題の簡単化のために，本研究ではそれぞれの基底は定常であるとする (一般に，アタック，サステイン，ディケイ，リリースなどでスペクトルパタンは異なることが多い)．各基底のパワーのみが時間変化すると考えれば，その時間変化 (音量や音長，音色) をアクティベーション行列 $\boldsymbol{U}$ とすることで，音楽スペクトログラムは次のように NMF と等価な問題として定式化できる．

音楽スペクトログラムを $\boldsymbol{Y} \in \mathbb{R}^{W\times T}$ とし，基底行列を $\boldsymbol{H} \in \mathbb{R}^{W\times K}$，アクティベーション行列を $\boldsymbol{U} \in \mathbb{R}^{K\times T}$ とすると，

$$\boldsymbol{Y} \approx \boldsymbol{H}\boldsymbol{U} \tag{1}$$

として観測スペクトログラムが低ランク行列積で近似できる．次に，単音モデルを用いてアクティベーション $\boldsymbol{U}$ を

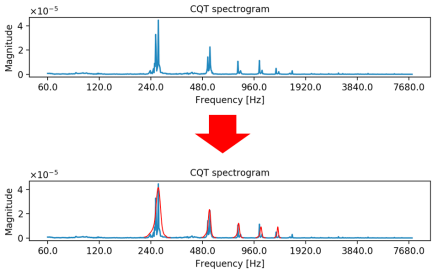

図 1 基底テンプレートの例．倍音に減衰したピークがたつよう混合正規分布で近似したスペクトルパタンを適用する．図は RWC Music Database [10] のピアノの C4(261.6Hz) のスペクトル構造を表す

さらに分解することを考える．単音のインデックスを r とし，それぞれの単音が持つエネルギーを V_r，ある時刻 (フレーム)t のアクティベーション形状を $\nu_{r,t}$ とすると，

$$U_{k,t} = V_r \nu_{r,t} \tag{2}$$

となる．また，表情付き演奏ピアノロールが楽譜演奏ピアノロールの伸縮により表現可能であるならば，楽譜演奏ピアノロールにより得られるオンセット時刻を事前情報として活用できる．したがって，$\nu_{r,t}$ を，理想的にはオンセット時刻にピークが立つような $O_{r,t}$ と，音長を表す時刻インデックス τ を用いた，パワーの減衰を表現する単音アクティベーション形状 $G_{r,\tau}$ とに分解し，その畳み込み表現 $\sum_\tau G_{r,\tau}O_{r,t-\tau}$ を用いて単音のアクティベーションの形状を表す．

以上より，音楽スペクトログラムは，

$$Y_{w,t} \approx Y'_{w,t} = \sum_{r,\tau} H_{w,\kappa_r} V_r G_{r,\tau} O_{r,t-\tau}$$
$$\forall_k \sum_w H_{w,k} = 1, \quad \forall_r \sum_\tau G_{r,\tau} = 1, \sum_t O_{r,t} = 1 \tag{3}$$

として表現できる．なお，κ_r は単音 r の時の音高 k を表す．以降，本章で述べるモデルを畳み込み単音重畳 (CSM) モデルと呼称する．

2.2 モデルの最適化

NMF では観測行列 $\boldsymbol{Y}$ と低ランク近似行列 $\boldsymbol{Y'}$ の距離最小化に基づく定式化を行う．本研究では距離基準として I-divergence を採用し，各パラメータに対して以下のような制約を与える．

$$S(\boldsymbol{H}) = \frac{\eta_H}{2}\sum_{w,k}||H_{w,k} - \hat{H}_{w,k}||^2$$
$$S(\boldsymbol{G}) = \mu_G \sum_{r,\tau}||G_{r,\tau}||^{p_g}$$
$$S(\boldsymbol{O}) = \mu_O \sum_{r,t}||O_{r,t}||^{p_o} + \frac{\eta_O}{2}\sum_{r,t}||O_{r,t} - \hat{O}_{r,t}||^2. \tag{4}$$

ここで，$\hat{\boldsymbol{H}}$ は事前学習や調波構造を与えることにより得られるテンプレート基底を表す．$\boldsymbol{G}, \boldsymbol{O}$ に対してはスパース

に誘導するための正則化項を与え，$\boldsymbol{O}$ にはさらに楽譜情報を参考に得られた推定オンセット分布 $\hat{\boldsymbol{O}}$ との二乗誤差制約を加えることで，楽譜情報から得られるオンセット分布を事前知識として利用している．なお，本研究ではテンプレート基底として図 1 で示すような調波構造を混合正規分布で近似したスペクトル構造を与え，推定オンセット分布は楽譜演奏 MIDI を WAVE 変換し得られたスペクトログラムとの DP マッチングにより得られた時刻を中心としてピークがたつ，左右対称に離散的な分布とした．

以上より、最適化すべき目的関数は，各種パラメータをまとめて $\boldsymbol{\Theta}$ として，

$$\begin{aligned}
minimize \quad & \mathcal{J}(\boldsymbol{\Theta}) := \sum_{w,t}\left(Y_{w,t}\log\frac{Y_{w,t}}{Y'_{w,t}} - Y_{w,t} + Y'_{w,t}\right) \\
& \qquad +S(\boldsymbol{H}) + S(\boldsymbol{G}) + S(\boldsymbol{O}) \\
where \quad & Y'_{w,t} = \sum_{r,\tau} H_{w,\kappa_r} V_r G_{r,\tau} O_{r,t-\tau} \\
s.t. \quad & \forall_{w,k}\ H_{w,k} \geq 0,\ \forall_r V_r \geq 0, \\
& \forall_{r,\tau}\ G_{r,\tau} \geq 0,\ \forall_{r,t} O_{r,t} \geq 0, \\
& \forall_k \sum_w H_{w,k} = 1, \\
& \forall_r \sum_\tau G_{r,\tau} = 1,\ \sum_t O_{r,t} = 1.
\end{aligned} \tag{5}$$

上式 $J(\boldsymbol{\Theta})$ はこのままでは解析的に解けないが，Jensen の不等式を用いた上限関数を設計することにより逐次最適化が可能となる．式 (5) の $\log Y'_{w,t}$ に対して，

$$\begin{aligned}
\log & \left(\sum_{r,\tau} H_{w,\kappa_r} V_r G_{r,\tau} O_{r,t-\tau}\right) \\
& \leq \sum_{r,\tau} \lambda_{w,t,r,\tau} \log\frac{H_{w,\kappa_r} V_r G_{r,\tau} O_{r,t-\tau}}{\lambda_{w,t,r,\tau}} \\
\lambda_{w,t,r,\tau} &= \frac{H_{w,\kappa_r} V_r G_{r,\tau} O_{r,t-\tau}}{\sum_{r,\tau} H_{w,\kappa_r} V_r G_{r,\tau} O_{r,t-\tau}}.
\end{aligned} \tag{6}$$

ただし，$\lambda_{w,t,r,\tau}$ は一つ前のステップで得られたパラメータを用いて計算する．また，スパース正則化項は，接線不等式を用いて，

$$\begin{aligned}
||G_{r,\tau}||^{p_g} &\leq p_g ||G'_{r,\tau}||^{p_g-1}(G_{r,\tau} - G'_{r,\tau}) + ||G'_{r,\tau}||^{p_g} \\
||O_{r,\tau}||^{p_o} &\leq p_o ||O'_{r,\tau}||^{p_o-1}(O_{r,\tau} - O'_{r,\tau}) + ||O'_{r,\tau}||^{p_o}
\end{aligned} \tag{7}$$

のように設計できる．

2.3 更新式

以上の上限関数を最小化するそれぞれのパラメータを求めることで，解析的に乗法更新式が得られる．紙面の都合上導出は省くが，更新式は次のように与えられる．

$$\begin{aligned}
H_{w,k} &= H'_{w,k} \cdot \frac{-b^H + \sqrt{(b^H)^2 - 4a^H c^H}}{2a^H} \\
G_{r,\tau} &= G'_{r,\tau} \cdot \frac{V_r \sum_t O_{r,t-\tau} \sum_w \frac{Y_{w,t} H_{w,\kappa_r}}{Y'_{w,t}}}{V_r \sum_t O_{r,t-\tau} + \mu_g p_g ||G'_{r,\tau}||^{p_g-1}} \\
O_{r,\tau} &= O'_{r,\tau} \cdot \frac{-b^O + \sqrt{(b^O)^2 - 4a^O c^O}}{2a^O} \\
V_r &= V'_r \cdot \frac{\sum_{w,t} \frac{Y_{w,t} H_{w,\kappa_r} \sum_\tau G_{r,\tau} O_{r,t-\tau}}{Y'_{w,t}}}{\sum_{t,\tau} G_{r,\tau} O_{r,t-\tau}} \\
a^H &= \eta_H \\
b^H &= \frac{\sum_t U_{k,t} - \eta_H \hat{H}_{w,k}}{H'_{w,k}} \\
c^H &= -\frac{1}{H'_{w,k}} \sum_t \frac{Y_{w,t} U_{k,t}}{Y'_{w,t}} \\
a^O &= \eta_O \\
b^O &= \frac{V_r \sum_t G_{r,t-\tau} + \mu_O p_o ||O'_{r,\tau}||^{p_o-1} - \eta_O \hat{O}_{r,\tau}}{O'_{r,\tau}} \\
c^O &= -\frac{V_r}{O'_{r,\tau}} \sum_{w,t} \frac{Y_{w,t} H_{w,k} G_{r,t-\tau}}{Y'_{w,t}}
\end{aligned} \tag{8}$$

二乗誤差制約を含む $H_{w,k}, O_{r,\tau}$ は二次方程式の解の形が現れてきているが，乗法更新則は保たれている．なお，H', G', O', V', Y' はそれぞれ一つ前のステップにおける値を示す．

3. 多重時間分解能 CSM-NMF

3.1 通常の NMF における課題

観測信号に対して STFT を行う際，解析フレーム長に起因する周波数分解能と時間分解能との不確定性原理のため，これらのトレードオフな問題が生じる．この問題に対し，[9] では高時間分解能と高周波数分解能な 2 つのスペクトログラムを利用し，並列に更新する並列 NMF が提案された．本章ではこれをさらに CSM モデルへと拡張した多重時間分解能 (MR)CSM-NMF の更新式を示す．

3.2 問題の定式化

短フレーム，長フレームにおける解析を $n = \{S, L\}$ とする．観測スペクトグラムを $\boldsymbol{Y}^n$ のように表すと，最適化すべき目的関数は，

$$\begin{aligned}
minimize \quad & \mathcal{J}(\boldsymbol{\Theta}) := \sum_{n,w_n,t_n}\left(Y^n_{w,t}\log\frac{Y^n_{w_n,t_n}}{Y'^n_{w_n,t_n}} - Y^n_{w_n,t_n} + Y'^n_{w_n,t_n}\right) \\
& \qquad +S(\boldsymbol{\Theta}) \\
where \quad & Y'^n_{w_n,t_n} = \sum_{r,\tau_n} H^n_{w_n,\kappa_r} V^n_r G^n_{r,\tau_n} O^n_{r,t_n-\tau_n} \\
s.t. \quad & \forall_{n,w_n,k}\ H^n_{w_n,k} \geq 0,\ \forall_{n,r} V^n_r \geq 0 \\
& \forall_{n,r,\tau_n}\ G^n_{r,\tau_n} \geq 0,\ \forall_{n,r,t_n} O^n_{r,t_n} \geq 0, \\
& \forall_k \sum_w H_{w,k} = 1, \\
& \forall_r \sum_\tau G_{r,\tau} = 1,\ \sum_t O_{r,t} = 1.
\end{aligned} \tag{9}$$

のように表せる．制約条件 $S(\boldsymbol{\Theta})$ は，前章で用いたテンプレート基底との二乗誤差制約，アクティベーションに関するパラメータのスパース正則化の他，長フレームにおける推定オンセット分布との二乗誤差制約，また短フレーム，長フレームの間の対応するフレーム，あるいは周波数ビンに対する類似制約を与える．結果として $S(\boldsymbol{\Theta})$ は，

$$
\begin{aligned}
S(\boldsymbol{H}^S) &= \frac{\mu_H^S}{2}\sum_{w_S,k}||H^S_{w_S,k}-\hat{H}^S_{w_S,k}||^2\\
S(\boldsymbol{H}^L) &= \frac{\mu_H^L}{2}\sum_{w_L,k}||H^L_{w_L,k}-\hat{H}^L_{w_L,k}||^2\\
S(\boldsymbol{H}^{SL}) &= \frac{\eta_H}{2}\sum_k\sum_{w_S}||b_{f,n_{w_S}}H^S_{w_S,k}-\sum_{w_L\in w_S}H^L_{w_L,k}||^2\\
S(\boldsymbol{G}^S) &= \mu_G^S\sum_{r,\tau_S}||G^S_{r,\tau_S}||^{p_{gs}}\\
S(\boldsymbol{G}^L) &= \mu_G^L\sum_{r,\tau_L}||G^L_{r,\tau_L}||^{p_{gl}}\\
S(\boldsymbol{G}^{SL}) &= \frac{\eta_G}{2}\sum_r\sum_{\tau_L}||b_{\tau,n_{\tau_L}}G^L_{r,\tau_L}-\sum_{\tau_S\in\tau_L}G^S_{r,\tau_S}||^2\\
S(\boldsymbol{O}^L) &= \mu_O^L\sum_{r,t_L}||O^L_{r,t_L}||^{p_{ol}}+\frac{\eta_O^L}{2}\sum_{r,t_L}||O^L_{r,t_L}-\hat{O}^L_{r,t_L}||^2\\
S(\boldsymbol{O}^S) &= \mu_O^S\sum_{r,t_S}||O^S_{r,t_S}||^{p_{os}}\\
S(\boldsymbol{O}^{SL}) &= \frac{\eta_O^S}{2}\sum_{r,t_S}||O^S_{r,t_S}-\sum_{t_L}O^L_{r,t_L}A_{t_L,t_S}||^2 \qquad (10)
\end{aligned}
$$

となる．ここで μ,η は各種制約に対する重み，b は対応する (共有する) ビン数，フレーム数を表す．すなわち，対応し合う基底 $\boldsymbol{H}^n$，オンセット分布 $\boldsymbol{O}^n$，形状分布 $\boldsymbol{G}^n$ の平均に対して二乗誤差制約をかけていることと等価である．また，A_{t_L,t_S} は長フレームにおけるオンセット分布 O^L_{r,t_L} を，対応するフレームに対して複製する変換行列である．

3.3 モデルの最適化

前章と同様に，Jensen の不等式及び接線不等式から上限関数を設計することで逐次更新可能となる．また，$(\sum_i x_i)^2$ となるような式に関しても同様に Jensen の不等式から，

$$
\begin{aligned}
\left(\sum_i x_i\right)^2 &\le \sum_i \frac{x_i^2}{\lambda_i}\\
\lambda_i &= \frac{x_i}{\sum_i x_i} \qquad (11)
\end{aligned}
$$

を利用できる．なお，導出の詳細は紙面の都合上割愛する．

3.4 更新式

3.4.1 $\boldsymbol{H^S_{w_S,k}}$

上限関数 (6),(7),(11) を用いて，式 (9) と式 (10) からなる評価関数 $\mathcal{J}(\boldsymbol{\Theta})$ を最小化する $H^S_{w_S,k}$ を求めると，

$$
H^S_{w_S,k} = H'^S_{w_S,k}\cdot\frac{-B_{H^S}+\sqrt{B^2_{H^S}-4(A_{H^S}C_{H^S})}}{2A_{H^S}} \qquad (12)
$$

ただし，

$$
\begin{cases}
A_{H^S} = \mu_H^S+\eta_H b^2_{f,n_{w_S}}\\
B_{H^S} = \frac{D_{H^S}}{H'^S_{w_S,k}}\\
C_{H^S} = -\frac{1}{H'^S_{w_S,k}}\sum_{t_S}Y^S_{w_S,t_S}\frac{U^S_{k,t_S}}{Y'^S_{w_S,t_S}}\\
D_{H^S} = \sum_{t_S}U^S_{k,t_S}-\mu_H^S\hat{H}_{w_S,k}\\
\qquad -\eta_H b_{f,n_{w_S}}\sum_{w_L\in w_S}H^L_{w_L,k}
\end{cases} \qquad (13)
$$

となる．更新式は二次方程式の解の公式で表される乗法更新となる．

3.4.2 $\boldsymbol{H^L_{w_L,k}}$

同様に，評価関数 $\mathcal{J}(\boldsymbol{\Theta})$ を最小化する $H^L_{w_L,k}$ を求めると，

$$
H^L_{w_L,k} = H'^L_{w_L,k}\cdot\frac{-B_{H^L}+\sqrt{B^2_{H^L}-4(A_{H^L}C_{H^L})}}{2A_{H^L}} \qquad (14)
$$

ただし，

$$
\begin{cases}
A_{H^L} = \mu_H^L H'^L_{w'_L,k}+\eta_H\sum_{w'_L\in w_S}H'^L_{w'_L,k}\\
B_{H^L} = \sum_{t_L}U^L_{k,t_L}-\mu_H^L\hat{H}^L_{w_L,k}\\
\qquad -\eta_H b_{f,n_{w_S}}H^S_{w_S,k}\\
C_{H^L} = -\sum_{t_L}Y^L_{w_L,t_L}\frac{U^L_{k,t_L}}{Y'^L_{w'_L,t_L}}
\end{cases} \qquad (15)
$$

となる．ここで，w_S は今着目している周波数インデックス w_L が含まれる短フレーム解析における周波数ビンを表す．

3.4.3 $\boldsymbol{G^S_{r,\tau_S}}$

同様に，評価関数 $\mathcal{J}(\boldsymbol{\Theta})$ を最小化する G^S_{r,τ_S} を求めると，

$$
G^S_{r,\tau_S} = G'^S_{r,\tau_S}\cdot\frac{-B_{G^S}+\sqrt{B^2_{G^S}-4A_{G^S}C_{G^S}}}{2A_{G^S}} \qquad (16)
$$

ただし，

$$
\begin{cases}
A_{G^S} = \eta_G\sum_{\tau'_S\in\tau_L}G'^S_{r,\tau'_S}\\
B_{G^S} = \sum_{t_S}V_r O^S_{r,t_S-\tau_S}+\mu_G^S p_{gs}||G'^S_{r,\tau_S}||^{p_{gs}-1}\\
\qquad -\eta_G b_{\tau,n_{\tau_L}}G^L_{r,\tau_L}\\
C_{G^S} = -\sum_{w_S,t_S}Y^S_{w_S,t_S}\frac{H^S_{w_S,\kappa_r}V_r O^S_{r,t_S-\tau_S}}{Y'^S_{w_S,t_S}}
\end{cases} \qquad (17)
$$

となる．ここで，t_L は今着目しているフレーム t_S と対応する長フレーム解析におけるフレームを表す．

3.4.4 $\boldsymbol{G^L_{r,\tau_L}}$

同様に，評価関数 $\mathcal{J}(\boldsymbol{\Theta})$ を最小化する G^L_{r,τ_L} を求めると，

$$
G^L_{r,\tau_L} = G'^L_{r,\tau_L}\cdot\frac{-B_{G^L}+\sqrt{B^2_{G^L}-4A_{G^L}C_{G^S}}}{2A_{G^L}} \qquad (18)
$$

ただし，

$$
\begin{cases}
A_{G^L} = \eta_G b^2_{\tau,n_{\tau_L}}\\
B_{G^L} = \frac{D_{G^L}}{G'^L_{r,\tau_L}}\\
C_{G^L} = -\frac{1}{G'^L_{r,\tau_L}}\sum_{w_L,t_L}Y^L_{w_L,t_L}\frac{H^L_{w_L,\kappa_r}V_r O^L_{r,t_L-\tau_L}}{Y'^L_{w_L,t_L}}\\
D_{G^L} = \sum_{t_L}V_r O_{r,t_L-\tau_L}+\mu_G^L p_{gl}||G'^L_{r,\tau}||^{p_{gl}-1}\\
\qquad -\eta_G b_{\tau,n_{\tau_L}}\sum_{\tau_S\in\tau_L}G^S_{r,\tau_S}
\end{cases} \qquad (19)
$$

となる．

3.4.5 O^S_{r,τ_S}

同様に，評価関数 $\mathcal{J}(\mathbf{\Theta})$ を最小化する O^S_{r,τ_S} を求めると，

$$O^S_{r,\tau_S} = O'^S_{r,\tau_S} \cdot \frac{-B_{O^S} + \sqrt{B^2_{O^S} - 4A_{O^S}C_{G^S}}}{2A_{O^S}} \quad (20)$$

ただし，

$$\begin{cases} A_{O^S} = \eta^S_O \\ B_{O^S} = \frac{D_{O^S}}{O'^S_{r,\tau_S}} \\ C_{O^S} = -\frac{1}{O'^S_{r,\tau_S}} \sum_{w_S,t_S} Y^S_{w_S,t_S} \frac{H^S_{w_S,\kappa_r} V_r G^S_{r,t_S-\tau_S}}{Y'^S_{w_S,t_S}} \\ D_{O^S} = \sum_{t_S} V_r G^S_{r,t_S-\tau_S} + \mu^S_O p_{os} ||O'^S_{r,\tau_S}||^{p_{os}-1} \\ \quad -\eta^S_O \hat{O}^S_{r,\tau_S} \end{cases} \quad (21)$$

となる．

3.4.6 O^L_{r,τ_L}

同様に，評価関数 $\mathcal{J}(\mathbf{\Theta})$ を最小化する O^L_{r,τ_L} を求めると，

$$O^L_{r,\tau_L} = O'^L_{r,\tau_L} \cdot \frac{-B_{O^L} + \sqrt{B^2_{O^L} - 4A_{O^L}C_{O^L}}}{2A_{O^L}} \quad (22)$$

ただし，

$$\begin{cases} A_{O^L} = \eta^L_O O'^L_{r,\tau_L} + \eta_O \hat{O''}^L_{r,t_L} \\ B_{O^L} = \sum_{t_L} V_r G_{r,t_L-\tau_L} + \mu^L_O p_{ol} ||O'^L_{r,\tau}||^{p_{ol}-1} \\ \quad -\eta^L_O \hat{O}^L_{r,\tau_L} - \eta_O \hat{O'}^L_{r,t_L} \\ C_{O^L} = -\sum_{w_L,t_L} Y^L_{w_L,t_L} \frac{H^L_{w_L,\kappa_r} V_r G^L_{r,t_L-\tau_L}}{Y'^L_{w_L,t_L}} \\ \hat{O'}^L_{r,t_L} = \sum_{t_S} O^S_{r,t_S} A_{t_L,t_S} \\ \hat{O''}^L_{r,t_L} = \sum_{t_S} \left(\sum_{t_L} O'^L_{r,t_L} A_{t_L,t_S} \right) A_{t_L,t_S} \end{cases} \quad (23)$$

である．

3.4.7 V^n_r

同様に，評価関数 $\mathcal{J}(\mathbf{\Theta})$ を最小化する V^n_r を求めると，

$$V^n_r = V'^n_r \cdot \sum_{w_n,t_n} \left(\frac{Y^n_{w_n,t_n} H^n_{w_n,\kappa_r} \frac{X^n_{r,t_n}}{Y'^n_{w_n,t_n}}}{H^n_{w_n,\kappa_r} X^n_{r,t_n}} \right) \quad (24)$$

となる．ただし，

$$X^n_{r,t_n} = \sum_{\tau_n} G^n_{r,\tau_n} O^n_{r,t_n-\tau_n} \quad (25)$$

である．

4. 評価実験

4.1 実験条件

これまでに述べた CSM-NMF 及び MRCSM-NMF に基づく更新式で求めたパラメータから，オンセット時刻及び音量の評価を行った．比較対象は楽譜演奏 MIDI を Classical archives [11] から，表情付き演奏 MIDI を International piano-e-competition [12] から，次の 2 曲 (i) Chopin, Ballade Op.52, No.4, (ii) Chopin, Prelude Op. 28, No. 24 について，それぞれ曲冒頭 20 秒程度を抽出し，Cubase を用いてサンプリングレート $16000Hz$ で WAVE 変換して解析した結果を用いた．解析フレーム長は，CSM モデルを 1024 のハーフオーバーラップとし，MRCSM モデルでは高周波数分解能を 4096，高時間分解能を 1024 として同様にハーフオーバーラップによる STFT を行った．

表 1 オンセット時刻の誤差パラメータと音量の相関

Music number	Onset time error	Correlation
(i) Op. 52, No. 4 (Signle)	$\mu = -0.51$, $\sigma^2 = 5.09$	$r = 0.80$
(i) Op. 52, No. 4 (Multi)	$\mu = -0.04$, $\sigma^2 = 11.1$	$r = 0.71$
(ii) Op. 28, No. 24 (Signle)	$\mu = 0.63$, $\sigma^2 = 1.13$	$r = 0.62$
(ii) Op. 28, No. 24 (Multi)	$\mu = 1,10$, $\sigma^2 = 7.32$	$r = 0.56$

また，各種正則化項の重み μ, η はいずれの場合も 1 とし，スパース正則化のための Lp ノルムも 1 とした．テンプレート基底は各音高における第 n 倍音が，

$$\frac{h(f_n)}{h(f_0)} = (n+1)^{-1.5}, \quad (26)$$

と減衰するよう設定した．推定オンセット分布は，楽譜演奏 WAVE と表情付き演奏 WAVE のスペクトログラムを FastDTW [13] によってマッチングし，対応するフレームにピークを与えた左右対称な離散分布を与えた．最大音長 (τ の最大) は 1.5 秒とした．

更新後のオンセット時刻は，推定して得られたオンセット分布及び形状パラメータ (MRCSM の場合は高時間分解能におけるパラメータ) を用いて式 (25) に基づき決定し，正解のオンセット時刻との差の分散がもっとも小さくなる値を閾値として設定した．また，音量は推定したオンセット時刻から 3 フレーム先までの間でもっとも大きい値となるアクティベーションの値 $U_{k,t}$ を抽出し，平均 0，分散 1 で正規化した値と，同様に正解ベロシティを正規化した値との相関によって評価した．

4.2 実験結果

表 1 は評価に用いた 2 曲について，それぞれのモデルにおける正解オンセット時刻 (フレーム) との差の平均及び分散と，音量に関しての相関係数である．Single は CSM-NMF，Multi は MRCSM-NMF を用いた結果である．両者ともに CSM-NMF の方が安定しており，MRCSM-NMF ではオンセット時刻ずれが大きくずれるものが CSM に比べて増えたために，全体として不安定な推定となってしまっている．また，ベロシティ (音量) の相関はいずれのモデルにおいても比較的高い値が得られている．CSM-NMF と比較して MRCSM-NMF で誤差分散が大きくなってしまった原因として考えられるのは，パラメータの増加に伴う推定誤差や，形状類似制約に起因する，同じ音高の単音同士の値をうまく分離できていないことが考えられる．実際に単音形状 $\sum_\tau G_{r,\tau} O_{r,t-\tau}$ ではなくアクティベーション $U_{k,t}$ を確認すると，図 2 に示すように，比較的安定した形状を得ることができていることがわかる．このことから，オンセット分

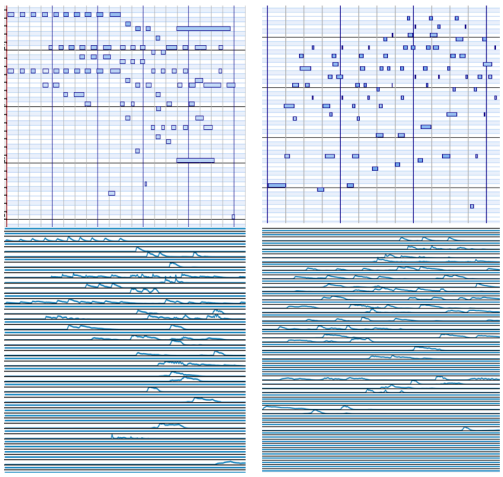

図 2 曲 (i)(左図) と (ii)(右図) の MRCSM-NMF による推定アクティベーション．上は正解 MIDI ピアノロールを示す

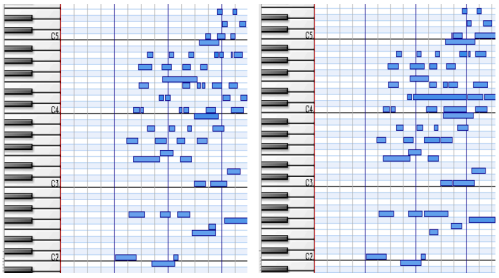

図 3 曲 (i) に対して CSM-NMF(左図) と MRCSM-NMF(右図) で推定したパラメータを用いてピアノロール生成を行った結果．

布 $O_{r,t}$ や形状パラメータ $G_{r,\tau}$ に対するスパース正則化パラメータの調整及び類似制約の緩和，あるいは類似制約だけではなく単調減少するような，形状に対する強い制約を付加することも有効であることも考えられる．事実予備的に複数のパラメータで実験した結果，スパース正則化の重みを強く，また類似制約を緩くすると，比較的安定した形状が得られやすい傾向が見られた．また，図 3 に (i) Chopin, Ballade Op.52, No.4 の冒頭部分に対し，CSM-NMF(左図) と MRCSM-NMF(右図) で推定したピアノロールを示す．

5. おわりに

本報告では，音楽音響信号に対する詳細解析を目標とし，楽譜情報を陽に活用する畳み込み単音重畳モデルの提案と，多重時間分解能モデルへの拡張を示した．CSM-NMF では，単音ごとのオンセット時刻や音量，音長 (あるいは音色) を同時に高精度に推定可能かつ有効であること，また，MRCSM-NMF においてもアクティベーションの推定は高精度に得られることを示した．

今後の課題として，パラメータ調整やオンセット推定閾値の動的決定，また音の立ち上がりに関するモデル化が検討する必要がある．その他応用として，事前情報を事前分布として設計するベイズモデルや，伸縮モデルを隠れマルコフモデルでモデル化することによる階層モデル化，さらには，異なる時間分解能により得られるアクティベーション形状の補間を利用した連続時間領域における形状関数の設計による，より詳細な時刻情報の解析などに取り組みたい．

謝辞

本研究は JSPS 科研費 17H00749 の支援を受けた．

参考文献

[1] C. Raphael, “Automatic transcription of piano music.” in *ISMIR*, 2002.
[2] M. Goto, “A real-time music-scene-description system: Predominant-f0 estimation for detecting melody and bass lines in real-world audio signals,” *Speech Communication*, vol. 43, no. 4, pp. 311–329, 2004.
[3] H. Katmeoka, T. Nishimoto, and S. Sagayama, “Separation of harmonic structures based on tied gaussian mixture model and information criterion for concurrent sounds,” in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 4. IEEE, 2004, pp. iv–iv.
[4] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, p. 788, 1999.
[5] M. D. Hoffman, D. M. Blei, and P. R. Cook, “Bayesian nonparametric matrix factorization for recorded music.” in *ICML*, 2010, pp. 439–446.
[6] E. Vincent, N. Bertin, and R. Badeau, “Adaptive harmonic spectral decomposition for multiple pitch estimation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 528–537, 2010.
[7] M. Nakano, J. Le Roux, H. Kameoka, T. Nakamura, N. Ono, and S. Sagayama, “Bayesian nonparametric spectrogram modeling based on infinite factorial infinite hidden markov model,” in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. IEEE, 2011, pp. 325–328.
[8] D. Liang and M. D. Hoffman, “Beta process non-negative matrix factorization with stochastic structured mean-field variational inference,” *arXiv preprint arXiv:1411.1804*, 2014.
[9] K. Ochiai, M. Nakano, N. Ono, and S. Sagayama, “Concurrent nonnegative matrix factorization using multi-resolution spectrograms for multipitch analysis of music signals,” *IPSJ SIG Technical Reports (MUS)*, vol. 2011, no. 5, pp. 1–6, 2011.
[10] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “Rwc music database : Database of copyright-cleared musical pieces and instrument sounds for research purposes,” *IPSJ Journal*, vol. 45, no. 3, pp. 728–738, 2004.
[11] Classical archives. [Online]. Available: https://www.classicalarchives.com/
[12] International piano-e-competition. [Online]. Available: http://www.piano-e-competition.com/
[13] S. Salvador and P. Chan, “Toward accurate dynamic time warping in linear time and space,” *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.