Vol.2014-MUS-103 No.18
2014/5/24

調波時間因子分解法に基づく事前情報付き多重音解析

四方 紘太郎[1]　高宗 典弘[1]　中村 友彦[1]　亀岡 弘和[1,2]

概要：本報告では、人間が音を知覚するメカニズムをもとにした知見を手がかりに，モノラル音楽音響信号の音高ごとの分離を目的とした調波時間構造化クラスタリングと非負値行列因子分解の利点を併せ持つ新しい音楽音響信号モデル，及びそれに基づく多重音解析手法である「調波時間因子分解法」を提案する。また、ビブラートや調などの音楽において特徴のある情報を補助情報として提案モデルに組み込んだ推論が可能なパラメータ推定アルゴリズムを提案する。

キーワード：多重音解析，音源分離，調波時間構造化クラスタリング，非負値行列因子分解，ウェーブレット変換，Product of Experts(PoE)，補助関数法

1. はじめに

複数の音源の信号が混合された観測信号から個々の音源に関する情報 (基本周波数，発音開始時刻，パワーなど) を抽出する処理である多重音解析は，音楽情報処理における重要課題の一つであり，自動採譜や音楽音響信号加工などの基礎技術となりうる．

マイクロホンアレイ入力からのブラインド音源分離では音源の空間的な手がかりを有効利用することができるが，モノラル音響信号を対象とした音源分離や多重音解析では空間的な手がかりに代わる何らかの仮定が必要である．聴覚情景分析における知見をヒントにしたアプローチである調波時間構造化クラスタリング (Harmonic-Temporal Clustering; HTC) [1], [2] は，人間が音をひとまとまりの音 (音脈) として知覚する要件 (調波性，連続性，同時性，同期性など) を時間周波数成分の局所的な制約として記述し，当該要件を満たすように観測信号の時間周波数成分を時間周波数平面上でクラスタリングしようというアイディアに基づいている．HTC ではこのアイディアを，各音脈に対応するスペクトログラムを拘束つき混合正規分布モデルとして記述し，それらを重畳したもので観測スペクトログラムにフィッティングするアプローチとして実現している．一方，モノラル音響信号を対象とした多重音解析手法として有効なアプローチとして近年注目されている非負値行列因子分解 (Non-negative Matrix Factorization; NMF) 法 [3] では，限られた種類の音高の楽音がそれぞれ異なるタイミングで繰り返し生起するという音楽特有の性質に着目し，限られた種類のスペクトルテンプレートの適当な重み付き和ですべての時刻の観測スペクトルを表せるはず，という仮定がベースとなっている．従って，観測スペクトログラムを非負値行列と見なし，これを二つの非負値行列の積（各スペクトルテンプレートを表す基底行列と，各荷重係数を要素にもつアクティベーション行列）に分解することにより観測スペクトログラムから各スペクトルテンプレートと各時刻におけるそれらの荷重係数を同時推定することができ，観測スペクトルを音高ごとのスペクトルに分解することが可能となるわけである．上記のとおり上述の二つのアプローチでは着目している手がかりが異なる．相対的には前者のアプローチでは局所的，後者のアプローチでは大域的な楽音の性質に基づいていると言え，いずれの性質も多重音解析を解決する上で本質的かつ有用な手がかりとなる．本稿では，上述の二つの性質を同時に取り入れた新しいスペクトログラムモデル，および当該モデルに基づく多重音解析手法「調波時間因子分解法」を提案する．

ところで，近年の音楽情報検索に関する国際会議や音楽情報検索に関わる各種タスクの国際コンテスト [4] において，調推定，和音推定，拍推定，オンセット推定などの手法の研究が急速に進展している．調，和音，拍，オンセットなどの情報が高い精度で得られるのであれば，多重音解析において極めて有用な補助情報となりうる．そこで本稿ではさらに，調推定，和音推定，拍推定，オンセット推定の各手法により得られる各種情報を前記提案モデルのパラメータ推論において補助情報として活用する枠組を提案する．音源分離において，ユーザのアシストにより分離精度を向上できるようにすることを目的としたユーザガイドつき音源分離 [5], [6] や，楽譜を補助情報とする楽譜ガイドつき音源分離 [7], [8] などの研究が進められているが，本稿で提案する枠組も補助情報ガイドつき音源分離の一種として位置づけられる．

以下，正規分布と Dirichlet 分布，Poisson 分布の確率密度関数を $\mathcal{N}$, Dir, Pois と表記する．

1 東京大学情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
2 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Comunication Science Labolatories, Nippon Telegraph and Telephone Coporation

2. 音楽スペクトログラムの確率モデル化

2.1 楽音信号のウェーブレット変換

本節ではまず，[1] に倣い，楽音信号のウェーブレット変換を導出する．

楽音の多くは，擬似周期信号 (局所的には周期的と見なせ，周期や調波成分のパワーが滑らかに時間変化する信号) と見なせる．そこで，音源 k の信号モデルとして n 次調波成分の瞬時位相が $n\theta_k(u)+\varphi_{k,n}$，瞬時振幅が $a_{k,n}(u)$ の擬似周期信号の解析信号表現

$$f_k(u)=\sum_{n=1}^{N}a_{k,n}(u)e^{j(n\theta_k(u)+\varphi_{k,n})} \tag{1}$$

を考える．ただし，u は時刻，$\varphi_{k,n}$ は初期位相である．紙面の都合上導出を省略するが，

$f_k(u)$ のウェーブレット変換 $W_k(x,t)$ を求める際に用いる基底関数 $\psi_{\alpha,t}(u)$ を次のように定義する．

$$\psi_{\alpha,t}(u)=\frac{1}{\sqrt{2\pi}\alpha}\psi\left(\frac{u-t}{\alpha}\right) \tag{2}$$

このとき，x は対数周波数，t は時間シフト変数，α はスケール，である．$\psi(u)$ は中心周波数が 1 のアナライジングウェーブレットである．また，$f_k(u)$ の連続ウェーブレット変換は次のように定義される．

$$W_k\left(\log\frac{1}{\alpha},t\right)=\langle f_k(u),\psi_{\alpha,t}(u)\rangle \tag{3}$$

このとき, $F_k(\omega)$ は $f_k(u)$ の Fourier 変換, $\Psi_{\alpha,t}(\omega)$ は $\psi_{\alpha,t}(u)$ の Fourier 変換である．また，Parseval の等式より，

$$\langle f_k(u),\psi_{\alpha,t}(u)\rangle=\langle F_k(\omega),\Psi_{\alpha,t}(\omega)\rangle \tag{4}$$

となる．ここで，$\psi(u)$ の Fourier 変換を $\Psi(\omega)$ とすると，(2)　の両辺の Fourier 変換は,

$$\Psi_{\alpha,t}(\omega)=\Psi(\alpha\omega)e^{-j\omega t} \tag{5}$$

となる．よって，(4)　より，

$$W_k\left(\log\frac{1}{\alpha},t\right)=\int_{-\infty}^{\infty}F_k(\omega)\Psi^*(\alpha\omega)e^{j\omega t}\mathrm{d}\omega \tag{6}$$

となる．ここで，楽音の擬似周期性の仮定から，$a_{k,n}(u)$ は近似的に定常なものとし，$\tilde{a}_{k,n}$ とおく．また，$\theta_k(u)$ も近似的に線形なものとみなし，$\tilde{\theta}_k(u)$ と近似する．このとき，$\dot{\tilde{\theta}}_k(u)=\tilde{\mu}_k u$ とおくと，$\tilde{\mu}_k$ は瞬時基本周波数 $\mu_k(u)$ の近似を意味する．この仮定を用いると，$f_k(u)$ の Fourier 変換は,

$$\begin{aligned}F_k(\omega)&=\frac{1}{\sqrt{2\pi}}\sum_{n=1}^{N}\int_{-\infty}^{\infty}\tilde{a}_{k,n}e^{j(n\tilde{\mu}_k u+\varphi_{k,n})}e^{-j\omega u}\mathrm{d}u \quad (7)\\&=\sqrt{2\pi}\sum_{n=1}^{N}\tilde{a}_{k,n}e^{j\varphi_{k,n}}\delta(\omega-n\tilde{\mu}_k)\end{aligned}$$

となる．よって，$f_k(u)$ のウェーブレット変換 $W_k(x,t)$ は，$x=\log\frac{1}{\alpha}$ の変数変換を施すと次のようになる．

$$W_k(x,t)=\sum_n a_{k,n}(t)\Psi^*(ne^{-x}\mu_k(t))e^{j(n\theta_k(t)+\varphi_{k,n})} \tag{8}$$

今，Ψ が次のような対数正規分布型の関数となるようなアナライジングウェーブレットを選ぶ.

$$\Psi(\omega)=\begin{cases}\exp\left(-\frac{(\ln\omega)^2}{4\sigma^2}\right) & (\omega>0)\\ 0 & (\omega\le 0)\end{cases} \tag{9}$$

また，$\Omega_k(t)=\ln\mu_k(t)$ と置くと，
$W_k(x,t)$ は

$$W_k(x,t)=\sum_n a_{k,n}(t)e^{-\frac{(x-\Omega_k(t)-\ln n)^2}{4\sigma^2}}e^{j(n\theta_k(t)+\varphi_{k,n})} \tag{10}$$

と表される．ただし，σ は対数正規分布のスケールパラメータに対応する定数である．ここで，n,n' $(n\neq n')$ の指数項の重なりがほとんどない (調波成分が互いに重ならない) と仮定できるならば，$|W_k(x,t)|^2$ は近似的に，

$$|W_k(x,t)|^2\simeq\sum_n|a_{k,n}(t)|^2e^{-\frac{(x-\Omega_k(t)-\ln n)^2}{2\sigma^2}} \tag{11}$$

と表すことができる．ここまでは HTC で採用された調波時間構造モデルと同一であり，混合正規分布モデル (Gaussian mixture model; GMM) と同形な関数となっていることが分かる．

2.2 音源スペクトログラムモデル

ここで，NMF におけるモデル化の考え方を上記モデルに取り入れる．NMF では，各楽音のスペクトルの形状は時不変で，スケールのみが時間変化すると仮定されるが，上記調波時間構造モデルにおいて，$|a_{k,n}(t)|^2$ を時刻 t に依らない変数と調波成分インデックス n に依らない変数の積の形

$$|a_{k,n}(t)|^2=w_{k,n}U_k(t)/\sqrt{2\pi}\sigma \tag{12}$$

に分解できるとすると，$|W_k(x,t)|^2$ は

$$|W_k(x,t)|^2=H_k(x,t)U_k(t) \tag{13}$$

$$H_k(x,t):=\sum_n\frac{w_{k,n}}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\Omega_k(t)-\ln n)^2}{2\sigma^2}} \tag{14}$$

と表される．上式の $H_k(x,t)$ は GMM 型の関数で表される音源 k の時刻 t におけるスペクトル形状を表しており，これにアクティベーション $U_k(t)$ が乗じられた形になっている．以下では w と U のスケールの任意性を除くため,

$$\sum_n w_{k,n}=1 \tag{15}$$

を仮定しておく．このとき，$w_{k,n}$ は調波成分のパワー比を表すパラメータとなる．既存モデルとの関連は 2.4 節 で詳しく述べるが，以上のモデルは HTC と NMF で用いられている音源スペクトログラムモデルの特長を併せ持っている．

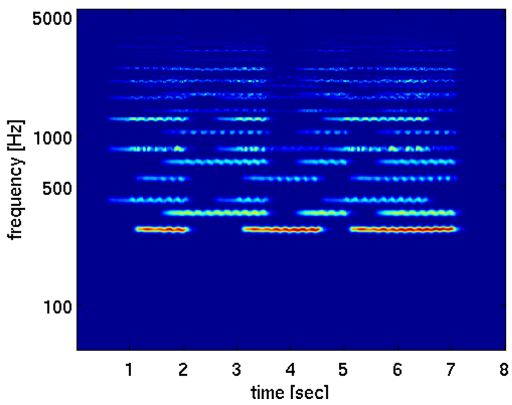

図 1 バイオリンのビブラート音のパワースペクトログラム

2.3 観測スペクトログラムの確率モデル

これまでは，スペクトログラムモデルを連続時間および連続周波数の関数として導いたが，実際に計算機で算出されるスペクトログラムは離散時刻および離散周波数ごとの値として得られる．そこで，以下では，等間隔に離散化された対数周波数を x_l $(l = 1, \dots, L)$，等間隔に離散化された時刻を t_i $(i = 1, \dots, I)$ とし，観測スペクトログラムを $Y(x_l, t_i)$ と表す．

HTC や NMF と同様にパワースペクトルの加法性を仮定すれば，音楽スペクトログラムは複数の音源スペクトログラム（式 (13)）を重畳したもの

$$X(x_l, t_i) = \sum_k H_k(x_l, t_i) U_k(t_i) \tag{16}$$

として表せる．このスペクトログラムモデル $X(x_l, t_i)$ には，実際の音源信号における式 (1) で定義した擬似周期性の仮定からの逸脱による誤差，調波間干渉を無視したことによる誤差，パワーの加法性の仮定に起因する誤差，背景雑音の存在に起因する誤差など，さまざまな要因の誤差が混在する．提案法の枠組では，これらの複合的な誤差要因を一つ一つ詳細にモデル化することはせず，まとめて一挙に確率的な現象と見なし，$Y(x_l, t_i)$ は

$$Y(x_l, t_i) \sim \mathrm{Pois}(Y(x_l, t_i); X(x_l, t_i)) \quad (\forall l, \forall i) \tag{17}$$

により生成されたものと仮定する．なお，この仮定の下で，$X(x_l, t_i)$ を変数と見なして最尤推定する問題は，スペクトル間の乖離度を I ダイバージェンスと呼ぶ歪み尺度を規準とした Y と X の最適フィッティング問題と等価となる．

2.4 従来モデルとの関連

本節では，先に提案したスペクトログラムモデル $X(x_l, t_i)$ と従来モデルとの関連について述べる．

まず，$H_k(x_l, t_i)$ に関し，式 (14) のようなパラメトリックな関数を仮定せずインデックス k, l, i ごとの $H_k(x_l, t_i)$ の値をパラメータと見なせば式 (16) は「可変基底 NMF」[9] において仮定されるスペクトログラムモデルと同一となる．また，$H_k(x_l, t_i)$ に対し時不変となるような拘束を置けば通常の NMF [3] において仮定されるスペクトログラムモデルと同一となる．さらに各 H_k に調波構造をなすスペクトルを仮定すれば，「調波 NMF」[10], [11] において仮定されるスペクトログラムモデルと同一となる．次に，式 (14) において，$\Omega_k(t)$ に対し時不変となるような拘束を置けば，[12], [13] のスペクトログラムモデルと同一となる．最後に，$U_k(t)$ を拘束つき GMM 型の関数で記述すれば，HTC[1], [2] において仮定されるスペクトログラムモデルと同一となる．以上より提案モデルは NMF と HTC の双方と親戚関係にあることが分かる．

3. 音楽事前情報の組み込み

3.1 事前分布としての音楽事前情報

音響信号から調や和音, 拍・オンセット時刻などの情報を推定するための手法の研究は近年急速に進展している [4]. 調や和音，拍・オンセット時刻などの情報は多重音解析においては有用な補助情報になりうるため，既存手法を用いてこれらの推定を前段で行い，その結果を補助情報として活用すれば高い精度で多重音解析を行える可能性がある．上記の前段処理では推定誤りを含むこともありえるが，補助情報としての信頼度を確率と捉えれば，各パラメータの事前確率として推論に組み込むことが可能である．

以上の補助情報を事前確率として表せれば，複数の推定結果の同時活用も可能である．例えば，調と和音の情報が与えられたときには，その調で出現しやすく，かつその和音で出現しやい音高が選ばれやすいはずである．この「かつ」に相当する演算は, 両方の条件を表す事前分布の積で表される．このように, 複数の条件を同時に成立させるには, 条件を表す事前分布の積として表現すれば良く，Products of Experts[14] の考え方が採用できる．

3.2 音楽事前情報の事前分布による設計

3.1 節の方針に従って，いくつかの音楽事前情報について事前分布を設計する．

3.2.1 基本周波数の事前分布

弦楽器や管楽器の楽器音は，1 つの音符を演奏していても基本周波数が変化しうる．例えば，バイオリンなどの奏法であるビブラートは，基本周波数がある音高に対応する基本周波数の周りで小刻みに振動しつつ，連続的に変化する (図. 1)．3.1 節で議論した通り，k 番目の音源スペクトログラムモデルの対数基本周波数 $\Omega_k(t_i)$ をビブラートで演奏されている音符の基本周波数とすると，この音高の基本周波数の周りでの変動を表す確率分布 q_g と基本周波数の連続的な変化を表す確率分布 q_s の積として事前分布を設計できる．

このような確率分布として，

$$q_g(\Omega_k(t_i)) = \mathcal{N}(\Omega_k(t_i); m_k, \nu_k^2) \tag{18}$$

$$q_c(\Omega_k(t_i)|\Omega_k(t_{i-1})) = \mathcal{N}(\Omega_k(t_i); \Omega_k(t_{i-1}), \tau_k^2) \tag{19}$$

を用いることができる．ここで，m_k, ν_k は k 番目の音源スペクトログラムモデルの音高に対応する対数基本周波数と対数周波数軸上での分散を表し，τ_k^2 は対数基本周波数の時間変化量の分散を表す．定性的に説明すると，q_c は対数基本周波数の軌跡が時間に関して滑らかであることを意味

し，q_g は時間に関係なく対数基本周波数が与えられた音高 m_k 周辺に存在することを意味する．q_g, q_c を用いて $\Omega_k(t)$ の事前分布は,

$$p(\Omega_k(t_i)|\Omega_k(t_{i-1})) \propto q_g(\Omega_k(t_i))^{\alpha_g} q_c(\Omega_k(t_i)|\Omega_k(t_{i-1}))^{\alpha_c} \tag{20}$$

と書ける．ここで，α_g, α_c は q_g, q_c の事前分布への寄与を調節するパラメータである．

3.2.2 調と和音の事前分布

調性のある楽曲においては，曲の部分ごとに調や和音が存在し，調と和音に従って出現する音高に偏りがある．これは，各音高に対応する音源スペクトログラムモデルのアクティベーション $U_k(t)$ の事前分布としてモデル化できる．また，従来よく用いられてきた事前分布も統合的に扱える．先行研究では，時間に関するスパース性を仮定することがあり，経験的に有効であることが知られている．

これら 2 つの条件を満たすような事前分布は, 楽曲全体の音量 C と時間方向に正規化された音量 $B_k(t_i)$ ($\sum_i B_k(t_i) = 1$)，音高方向に正規化された音量 A_k ($\sum_k A_k = 1$) を用いて,

$$U_k(t_i) = CA_kB_k(t_i), \tag{21}$$

$$\boldsymbol{A} := [A_1, \ldots, A_K]^\top \sim \mathrm{Dir}(\boldsymbol{A}; \boldsymbol{\beta}) \tag{22}$$

$$\boldsymbol{B}_k := [B_k(t_1), \ldots, B_k(t_I)]^\top \sim \mathrm{Dir}(\boldsymbol{B}_k; \boldsymbol{\gamma}_k) \tag{23}$$

とモデル化できる．このように，調や和音による各音高の生起しやすさは，$\boldsymbol{A}$ の事前分布のハイパーパラメーター $\boldsymbol{\beta}$ を適切に設定することにより反映することが出来る．また，時間に関するスパース性も $\boldsymbol{B}_k$ の事前分布としてハイパーパラメータ $\boldsymbol{\gamma}_k$ を適切に設定し，反映することができる．

4. 事後確率最大化によるパラメータ推定アルゴリズム

4.1 目的関数

2，3 章の議論に基づき，音楽パワースペクトログラム $Y := \{Y(x_l, t_i)\}_{l,i}$ が与えらたときのパラメータ推定法を考える．事前分布を含む最尤推定は MAP 推定とよばれ，パラメータ $\Theta = \{w_{k,n}, \Omega_k(t_i), A_k, B_k(t_i), C\}_i$ を用いて,

$$\underset{\Theta}{\mathrm{argmax}} \ln p(\Theta|Y) = \underset{\Theta}{\mathrm{argmax}} (\ln p(Y|\Theta) + \ln p(\Theta)) \tag{24}$$

となる事後確率 $\ln p(\Theta|Y)$ を最大化する Θ を推定する．ここで，$\ln p(Y|\Theta)$ は尤度，$p(\Theta)$ は事前分布を表し,

$$\begin{aligned}&\ln p(Y|\Theta)\\ &=_c \sum_{i,l}\left(Y(x_l,t_i)\ln\left(\sum_k H_k(x_l,t_i)U_k(t)\right)\right.\\ &\qquad\left.-\sum_k H_k(x_l,t_i)U_k(t)\right)\end{aligned} \tag{25}$$

$$\begin{aligned}&\ln p(\Theta)\\ &=_c \sum_{k,i}(\alpha_g \ln q_g(\Omega_k(t_i)) + \alpha_c \ln q_s(\Omega_k(t_i)|\Omega_k(t_{i-1})))\\ &\qquad+\sum_i \ln \mathrm{Dir}(\boldsymbol{A};\boldsymbol{\beta}) + \sum_k \ln \mathrm{Dir}(\boldsymbol{B}_k;\boldsymbol{\gamma}_k)\end{aligned} \tag{26}$$

と書ける．ここで，$=_c$ は定数を除いて一致することを表す．ところで，式 (25) の第 2 項は次のように書ける．

$$\begin{aligned}&\sum_{i,l,k} H_k(x_l,t_i)U_k(t_i)\\ &=\sum_{l,i,k,n} U_k(t_i)\frac{w_{k,n}}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x_l-\Omega_k(t)-\ln n)^2}{2\sigma^2}\right)\end{aligned} \tag{27}$$

ここで,

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\Omega_k(t_i)-\ln n)^2}{2\sigma^2}\right)\mathrm{d}x = 1 \tag{28}$$

より，区分求積法による近似を行い，正規分布の縁の値が数値的に無視できるものとすると,

$$\begin{aligned}&\sum_{l,n} U_k(t_i)\frac{w_{k,n}}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x_l-\Omega_k(t_i)-\ln n)^2}{2\sigma^2}\right)\\ &\approx \frac{U_k(t_i)}{X_0}\end{aligned} \tag{29}$$

と近似できる．ただし，X_0 は離散対数周波数の間隔を示す．(24) の右辺を $\arg\max_\Theta J(\Theta)$ とすると，$J(\Theta)$ を最大化するような Θ の更新則を求めれば良い．しかし，(25) には対数関数の中に k, n に関する和が存在するため，閉形式の更新則を求めることは容易ではない.

この場合には，補助関数法 [15] を用いることにより閉形式の更新則を導くことができる．補助関数法は，補助変数を用いて $J(\Theta)$ の下界（補助関数）を作り，その補助関数を補助変数と Θ について交互に最大化することにより $J(\Theta)$ を単調増加させる．

(25) の問題となる部分は，対数関数が凹関数であるため Jensen の不等式を用いて,

$$\begin{aligned}&\sum_{i,l} Y(x_l,t_i)\ln\left(\sum_{k,n} w_{k,n}\phi_{k,n}(x_l,t_i)U_k(t)\right)\\ &\geq \sum_{i,l} Y(x_l,t_i)\sum_{k,n}\lambda_{i,l,k,n}U_k(t_i)\ln\frac{w_{k,n}\phi_{k,n}(x_l,t_i)}{\lambda_{i,l,k,n}}\end{aligned} \tag{30}$$

と補助変数 $\lambda_{i,l,k,n} \in [0,1]$ を使用して下界を求めることができる．ここで，$\forall i, l, \sum_{k,n} \lambda_{i,l,k,n} = 1$ である．等号成立条件は,

$$\lambda_{i,l,k} = \frac{U_k(t_i)w_{k,n}\phi_{k,n}(x_l,t_i)}{\sum_{n,k} U_k(t_i)w_{k,n}\phi_{k,n}(x_l,t_i)}, \tag{31}$$

である．したがって，$J(\Theta)$ の補助関数は,

$$\begin{aligned}&J^+(\Theta, \{\lambda_{i,l,k,n}\}_{k,n})\\ &=_c \sum_{i,l} Y(x_l,t_i)\sum_{k,n}\lambda_{i,l,k,n}U_k(t_i)\ln\frac{w_{k,n}\phi_{k,n}(x_l,t_i)}{\lambda_{i,l,k,n}}\\ &\quad -\frac{U_k(t_i)}{X_0}\\ &\quad +\sum_{k,i}(\alpha_g \ln q_g(\Omega_k(t_i)) + \alpha_c \ln q_s(\Omega_k(t_i)|\Omega_k(t_{i-1})))\\ &\quad +\sum_i \ln \mathrm{Dir}(\boldsymbol{A};\boldsymbol{\beta}) + \sum_k \ln \mathrm{Dir}(\boldsymbol{B}_k;\boldsymbol{\gamma}_k)\end{aligned} \tag{32}$$

と書ける.

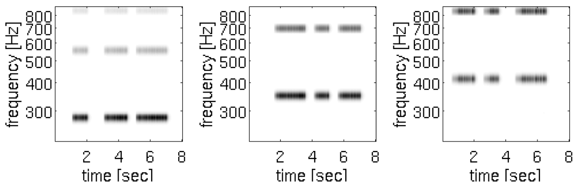

図 2 *I*-ダイバージェンス基準 NMF による基底スペクトログラムの推定結果 (基底数 3)

4.2 パラメータの更新則

補助関数法では，パラメータの更新と補助変数の更新を繰り返して最適化を行う．補助変数の更新は等号成立条件 (31) にしたがって更新すれば良い．パラメータの更新則は，補助関数 $J^{+}(\Theta, \{\lambda_{i,l,k}, \zeta_{i,l,k,n}\}_{k,n})$ を各パラメータで偏微分した値が 0 となるように求めればよい．C は音楽パワースペクトログラム Y から一意に決まり，

$$C = X_0 \sum_{i,l} Y(x_l, t_i) \tag{33}$$

である．パラメータの更新則はそれぞれ

$$w_{k,n} \leftarrow \frac{\sum_{l,i} Y(x_l,t_i)\lambda_{i,l,k,n}}{\sum_{l,i,n} Y(x_l,t_i)\lambda_{i,l,k,n}} \tag{34}$$

$$\begin{aligned} \mathbf{\Omega}_k \leftarrow & \left(\frac{\alpha_c}{\tau^2} D^{\top} D + \frac{\alpha_g}{\nu^2} E_I + (\mathrm{diag}\ \mathbf{p}_{l,k,n})\right)^{-1} \\ & \left(\frac{m_k \alpha_c}{\nu^2} \mathbf{1}_I + \sum_{l,n} (x_l - \ln n)\ \mathbf{p}_{l,k,n}\right) \end{aligned} \tag{35}$$

$$A_k \leftarrow \frac{\sum_{l,i,n} Y(x_l,t_i)\lambda_{i,l,k,n} + \beta_k - 1}{\sum_k (\sum_{l,i,n} Y(x_l,t_i)\lambda_{i,l,k,n} + \beta_k - 1)} \tag{36}$$

$$B_k(t_i) \leftarrow \frac{\sum_{l,n} Y(x_l,t_i)\lambda_{i,l,k,n} + \gamma_{i,k} - 1}{\sum_t (\sum_{l,n} Y(x_l,t_i)\lambda_{i,l,k,n} + \gamma_{i,k} - 1)} \tag{37}$$

と導出できる．ここで，diag はベクトルを行列の対角要素に順に並べる演算を表し，$\mathbf{1}_I$ は I 次元で要素が全て 1 のベクトル，E_I は I 次元の単位行列である．$\mathbf{\Omega}_k$ は

$$\mathbf{\Omega}_k := [\Omega_k(t_1), \ldots, \Omega_k(t_I)]^{\top} \tag{38}$$

で定義され，$\mathbf{p}_{l,k,n}$ は

$$p_{l,k,n,i} = \frac{Y(x_l,t_i)\lambda_{i,l,k,n}}{\sigma^2} \tag{39}$$

を用いて，$\mathbf{p}_{l,k,n} := [p_{l,k,n,1}, \ldots, p_{l,k,n,I}]^{\top}$ と書け，$I \times I$ 行列 D は

$$D = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ 0 & & & -1 & 1 \end{pmatrix} \tag{40}$$

である．これを調波時間因子分解法と名付ける．

5. 多重音解析動作実験

本研究において提案した手法を MATLAB で実装し，多重音解析動作実験を行った．本章ではそれに対し定性的な考察を試みた．

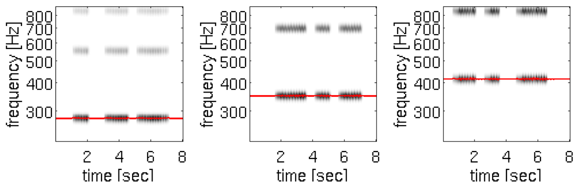

図 3 D♭,F,A♭ それぞれの音源スペクトログラムモデルと基本周波数の時間軌跡推定結果 (調に関する音楽事前情報なし)．赤線は各時刻の $\Omega_k(t_i)$ の推定値を表す．

5.1 ビブラートの基本周波数推定実験

提案モデルは基本周波数の時間変動を考慮しており，ビブラート音の基本周波数の推移を追うことが可能であると期待される．そこで，RWC 楽器音データベース [16] を用いて，D♭，F，A♭ のバイオリン音源のビブラート音を合成した音源（サンプリング周波数 16 kHz）を作成し，基本周波数の推定に関し，*I*-ダイバージェンス基準の NMF(基底数 3, パラメータ更新回数 100 回) との比較実験を行った．

スペクトログラムを音響信号から得る際に, 高速近似連続ウェーブレット変換 [17] を用いた．この変換を用いると, スペクトログラムの時間シフトはある程度自由に設定でき, 本実験では $t_i - t_{i-1} = 16$ ms とした．アナライジングウェーブレットは，対数正規分布型のウェーブレット [1] を用いて $x_1 = \ln(55), x_l - x_{l-1} = \ln(2)/120$ となるようにスケール値を設定した．提案法のパラメータは $(N, K, \tau, \nu, \sigma, \alpha_g, \alpha_c) = (8, 73, 2 \times 10^4, 1.5, \times 10^4, 0.02, 1, 1)$，$\boldsymbol{\gamma}_k = (1 - 3.96 \times 10^{-6})\mathbf{1}_I$，$\boldsymbol{\beta} = (1 - 2.4 \times 10^{-3})\mathbf{1}_K$ とし，パラメータ更新回数は 10 回とした．

観測 (図. 1) で見られたビブラートは，図. 2 のように NMF のモデルスペクトログラムで平坦な形になり，ビブラートの基本周波数の時間的な変化を捉えられていないことがわかる．それに比べ，提案手法では図. 3 のように基本周波数パラメータが時間領域でビブラート時の基本周波数の推移と同様の軌跡を描いていることが確認できる．

5.2 音楽事前情報の効果確認の実験

次に，調の情報を事前分布として取り入れた場合，調に対応する音階外の音が抑制されるかを確認する実験を行った．ここで調は D♭ であると仮定し，音階外である D の音に対するアクティベーションが抑制されているかを確認した．音階内の音について $\beta = 1 - 2.4 \times 10^{-3}$ とし，音階外の音については $\beta = 1 - 3.0 \times 10^{-3}$ とした．その他のパラメータは 5.1 節と同じである．

図. 4 右が示す通り，調の音楽事前情報を反映することによって，調の音楽事前情報を用いていない図 4 左に比べて音階外である D の音高 (赤線) のアクティベーションが抑制された．したがって，音楽事前情報に音高の誤推定を抑制する効果があることを確認できた．

6. 結論

本研究では，NMF と HTC の性質を同時に取り入れた新たな音源スペクトログラムモデルを提案し，そのモデルに基づく多重音解析手法である調波時間因子分解法を提案

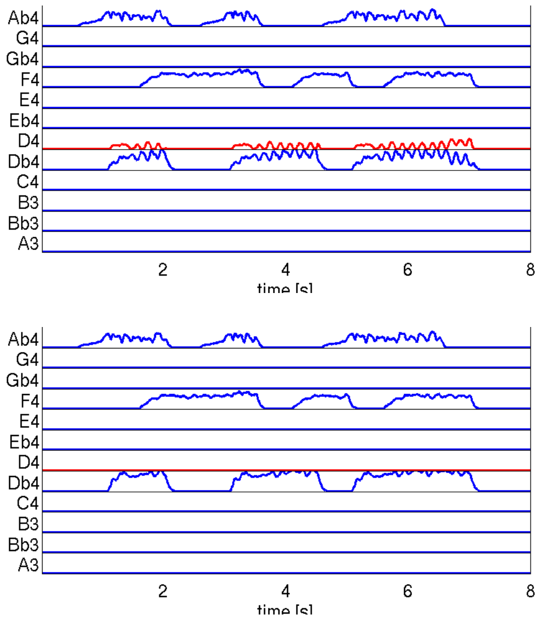

図 4 音楽事前情報なし (上), あり (下) のときの各音高のアクティベーション推定結果

した．また，音楽事前情報を Product of Experts の形で事前分布としてモデルに導入できることを示し，事後確率を最大化するような推定アルゴリズムを導出した．基本周波数の時間変化に関しては，ビブラート音に対する基本周波数推定実験から，基本周波数パラメータが時間領域で基本周波数の推移に追従できることを確認した．また，音楽事前情報による多重音解析性能への効果を確認するために，調を音楽事前情報として用いた実験を行い，音階外の音高の誤推定が抑制されることを確認した．

今後は，提案手法の定量的な評価を行い，音楽事前情報を取り入れた高精度な多重音解析ソフトウェアを開発することが課題である．

参考文献

[1] Hirokazu Kameoka, "Statistical approach to multipitch analysis," Ph.D. thesis, University of Tokyo, 2007.
[2] Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," IEEE Transactions on Audio, Speech and Language Processing, Vol. 15, No. 3, pp. 982–994, Mar. 2007.
[3] Paris Smaragdis, and Judith C. Brown, "Non-Negative Matrix Factorization for Polyphonic Music Transcription, " In *Proc.* the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, USA, 2003.
[4] `http://www.music-ir.org/mirex/wiki/MIREX\_HOME`
[5] Paris. Smaragdis, and Gautham J. Mysore, "Separation by "humming": Userguided sound extraction from monophonic mixtures," In *Proc.* the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 69–72, 2009.
[6] Alexey Ozerov, Cédric Févotte, Raphaël Blouet, and Jean-Louis Durrieu, "Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation," In *Proc.* the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 257–260, 2011.
[7] Romain Hennequin, Bertrand David, and Roland Badeau, "Score informed audio source separation using a parametric model of non-negative spectrogram," In *Proc.* the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 45–48, 2011.
[8] Umut Simsekli, A. Taylan Cemgil, "Score guided musical source separation using generalized coupled tensor factorization." In *Proc.* European Signal Processing Conference, pp. 2639–2643, 2012.
[9] Masahiro Nakano, Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono, Shigeki Sagayama, "Infinite-State Spectrum Model for Music Signal Analysis," In *Proc.* the IEEE International Conference on Acousitcs, Speech, and Signal Processing, 2011.
[10] Stanislaw A. Raczyński, Nobutaka Ono, and Shigeki Sagayama, "Multipitch analysis with harmonic nonnegative matrix approximation," In *Proc.* The International society for Music Information Retrieval, pp. 381–386, 2007.
[11] Emmanuel Vincent, Nancy Bertin, and Roland Badeau, "Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription," In *Proc.* The IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 109–112, 2008.
[12] Kazuyoshi Yoshii, and Masataka Goto, "Infinite Latent Harmonic Allocation: A Nonparametric Bayesian Approach to Multipitch Analysis" , In *Proc.* The International society for Music Information Retrieval, pp. 309–314, 2010.
[13] 阪上 大地, 大塚 琢馬, 糸山 克寿, 奥乃 博, "非負値調波時間構造因子分解法に基づく音楽音響信号の多重基本周波数解析," 情報処理学会第 75 回全国大会, 4T-8, 2013.
[14] Geoffrey E. Hinton, "Training products of experts by minimizing contrastive divergence." *Neural computation*, 14(8), pp. 1771–1800, 2002.
[15] 亀岡弘和, 後藤真孝, 嵯峨山茂樹, "スペクトル制御エンベロープによる混合音中の周期および非周期成分の選択的イコライザ," 情報処理学会研究報告, 2006-MUS-66-13, pp. 77–84, Aug. 2006.
[16] Masataka Goto, "Development of the RWC Music Database," In *Proc. of* the 18th International Congress on Acoustics (ICA 2004), pp. I-553–556, April 2004.
[17] 亀岡弘和, 田原鉄也, 西本卓也, 嵯峨山茂樹, "信号処理方法及び装置," 特開 2008-281898, 11, 2008.