Vol.2012-MUS-94 No.27
Vol.2012-SLP-90 No.27
2012/2/4

調波・非調波・音色構造因子分解による音響信号分析と音源分離インターフェースへの応用

安良岡　直希[†1]　　奥乃　博[†2]

本稿では，多重奏音楽音響信号の振幅スペクトログラムを，J 種類の音色構造と K 種類の調波構造の組み合わせ，及び L 種類の異なる音色を持つ非調波構造，の 3 要素に分解する新しい音響信号分析法について述べる．調波構造を調波ガウス関数列で，音色構造を全極型伝達関数でそれぞれモデル化し，その和で構成されるスペクトログラムモデルのパラメータ (音高，音量，音色に対応) を補助関数法を用いて一挙に推定する．また，推定結果を用いて楽器パートごとの音量操作を行う試作 GUI をいくつかの動作例とともに紹介する．評価実験では非負値行列因子分解との比較と行い，提案法の有効性を示す．

Musical Audio Signal Modeling for Joint Estimation of Haromnic, Inharmonic, and Timbral Structure and its Application to Source Sepatation

NAOKI YASURAOKA[†1] and HIROSHI G. OKUNO [†2]

This paper presents a new method for polyphonic music spectrogram modeling, The method decomposes polyphonic spectrogram into three types of factors: combination of J timbral structures and K harmonic structures, and L inharmonic timbral structures. Harmonic Gaussian functions and an all-pole transfer function are introduced for representing harmonic structure and timbral structure, respectively. The auxiliary function method is used for estimating the model parameters, which consists of fundamental frequencies, all-pole coefficients and volumes of each element. A GUI designed for musical source separation with some separation examples is also introduced. Experimental result shows the proposed method separates each musical part more accurately in comparison with another one based on nonnegative matrix factorization.

1. はじめに

市販 CD のような多重奏の音響信号から、個々の楽器の種類や各単音など『パーツ』の情報を推定するという課題は、楽曲の内容に基づく検索推薦、自動採譜、音源分離、楽曲再加工など幅広い応用に共通するものである。近年の動向として、一個人では試聴しきれない数の楽曲がインターネットから簡単に参照・購入できること、専門知識を持たない一般の人々が作曲・創作活動を行い、さらに二次・三次創作へと波及する事例が急増していること、などが挙げられ、上記『パーツ』推定の技術の需要は今後より一層高まると予想される。

『パーツ』推定の基本は人間が感じる音の単位に沿って音響信号を分解することである．人間は複数の音から個々の『パーツ』を聞き分けるために，主に次の情報を活用する．

(1)　調波的な音か非調波的な音か
(2)　調波的な音なら，その音高
(3)　音色 (大局的な周波数特性)

コンピュータを用いた楽曲分析処理も，上記指標に基づいて音響信号を分解できれば，混合音響信号中の特定楽器だけに対し，音量を調節する，音色を変える，エフェクトを付加する，など柔軟な楽曲加工が可能になると期待される．従来より様々な混合音分析法が報告されているが，この 3 要素を一挙に推定する手法は未だ研究段階にあると言える[1)–4)]．

本稿では，上記の『パーツ』に沿って音楽音響信号の振幅スペクトログラムを分解する新しい手法『調波・非調波・音色構造因子分解 (Harmonic-Inharmonic-Timbral Factorization: HITF)』を報告する．HITF では，混合音響信号の振幅スペクトログラムを，J 種類の音色構造と K 種類の調波構造の組み合わせ，及び L 種類の異なる音色を持つ非調波構造，の和によってモデル化する．入力された混合音響信号と，上記モデルが示す振幅スペクトログラムが似た形状となるように，このモデルのパラメータ：各要素信号の各時刻毎の音量，K 個の調波構造の各時刻毎の基本周波数，$J+L$ 個の音色構造関数の係数，の値を一挙に推定する．推定結果は，**図 1** のように，J 種類の調波音色の各ピアノロール，L 個の非調波音色の音量軌跡として可視化できる．ここから，各要素信号の音量を変更したりと，楽譜情報として出力したり，様々な音楽情報処理へと展開できると考えられる．

†1 ヤマハ株式会社
Yamaha Corporation
†2 京都大学 大学院情報学研究科
Graduate School of Informatics, Kyoto University

Vol.2012-MUS-94 No.27
Vol.2012-SLP-90 No.27
2012/2/4

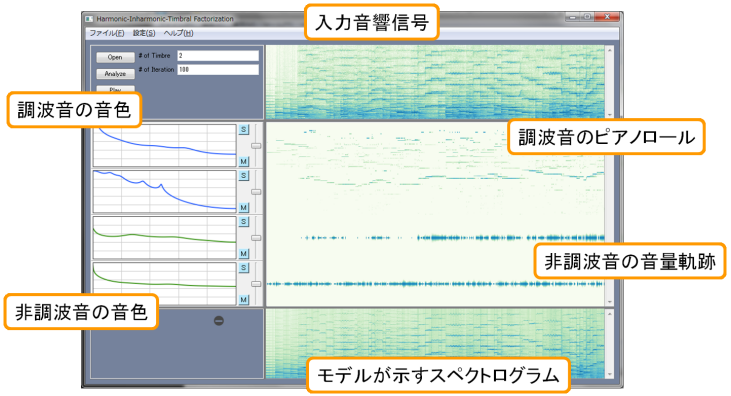

図 1 HITF の動作 GUI インターフェース：ピアノとバイオリン演奏の分析結果.

2. HITF に基づく音源分離

2.1 問題設定

本研究は幅広い音響信号加工技術を目指したものであるが，本稿では音源分離：音楽音響信号をユーザ所望の要素ごとに分離する問題に限定して議論する．分離の単位は，楽器ごとや，個々のノートごとなどいくつか方針が考えられるが，本稿では調波音と非調波音に分離し，その各々が数個の音色に分類され，さらに調波音は音高ごとに分離されることを目指す．つまり，**図 2** のような混合過程のモデルを想定し，その逆処理を行う．扱う音楽音響信号はモノラルとする．システムへの入力は音楽音響信号のみで，楽譜等の事前情報は持たない．

本稿で述べる音源分離法は時間周波数領域での処理に基づく．すなわち，入力信号を Short Time Fourier Transform (STFT) し得られるスペクトログラム $Y_{n,f}\phi_{n,f}$ を要素信号のスペクトログラムに分ける．ここで，スペクトログラムの振幅成分が $Y_{n,f}$ であり，位相成分が $\phi_{n,f}$ とする．n と f はそれぞれ時間フレームと周波数ビンを指すインデックスである．分離問題を簡略化するため，本音源分離法は振幅スペクトログラムの分配に基づいている．すなわち，$Y_{n,f}$ のモデル $X_{n,f}$ はいくつかの要素振幅スペクトログラム $X_{j,n,f}$ の和からなり，

$$Y_{n,f} \simeq X_{n,f} := \sum_j X_{j,n,f} \tag{1}$$

なる関係を持つように $X_{j,n,f}$ を推定する問題として定式化する ($\simeq$ は近似，$:=$ は定義を示す)．複数の音響信号の混合は「複素」スペクトログラム上での加算なので，非負値を分配する式 (1) の方法は要素信号の混合過程の逆演算にはなり得ないが，多くの音源分離法がこの方針に基づいており，それらは妥当な結果を得ている[1),3)]．また，振幅スペクトログラムの分配に基づく音源分離法を複素領域に拡張した例もある[5)]．

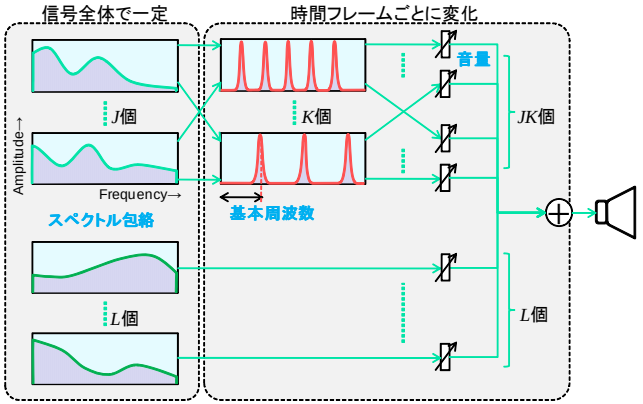

図 2 HITF が想定する要素信号構成

2.2 HITF の混合過程モデルの説明

HITF では，図 2 の考えに沿って振幅スペクトログラムのモデルを定義する．このモデルでは，音高を明確に感じる音は調波構造を持ち，その音色はスペクトル包絡で説明できると仮定している．時刻 n の振幅スペクトルは，基本周波数 μ_n^k を持つ K 個の調波構造 $G_{n,f}^k$ と J 個のスペクトル包絡関数 $1/|A_f^j|$ の組み合わせ，及び L 個の非調波音用のスペクトル包絡関数 $1/|B_f^l|$ の各要素スペクトルの和で構成される．時刻 n における，j,k 番目の調波成分の音量を $H_n^{j,k}$，l 番目の非調波成分の音量を I_n^l とすると，HITF のモデルは具体的に

$$X_{n,f} := \sum_{j,k} \frac{G_{n,f}^k}{|A_f^j|} H_n^{j,k} + \sum_l \frac{1}{|B_f^l|} I_n^l \tag{2}$$

と書き表せる．

調波構造は具体的には，基本周波数パラメータ μ_n^k に応じて周波数方向に伸縮するような，等間隔に並ぶガウス関数列で定義する．

$$G_{n,f}^k = \sum_h \exp\left[-\frac{(\hat{f} - h\mu_n^k)^2}{2\sigma^2}\right] \tag{3}$$

Vol.2012-MUS-94 No.27
Vol.2012-SLP-90 No.27
2012/2/4

ここで，h は倍音のインデックスであり，$\hat{f}$ は周波数ビン f に対応する周波数 (Hz) である．σ^2 は周波数方向の広がりであり，主に STFT 条件のみに影響されるパラメータであるので音響信号全体で単一の値を設定する．この調波構造モデルは基本周波数推定法ハーモニッククラスタリング[2] で用いられるものであり，各時刻，各単音ごとにパラメータ μ_n^k を可変にすればこの関数が周波数方向に伸縮しビブラートのような音高変化を正しく推定できる．

音色はスペクトル包絡構造と関係が深いことが知られており，本稿では包絡構造をパラメータ α_p^j, β_q^l を持つ全極型伝達関数でモデル化する．

$$\frac{1}{|A_f^j|} := \frac{1}{\left|1 - \sum_p \alpha_p^j \mathrm{e}^{-i\check{f}p}\right|}, \qquad \frac{1}{|B_f^l|} := \frac{1}{\left|1 - \sum_p \beta_q^l \mathrm{e}^{-i\check{f}p}\right|} \tag{4}$$

ここで, i は虚数単位であり，$\check{f} = 2\pi f/(F-1)$ は, 正規化角周波数である (F は正の周波数ビンの個数)．全極型係数の個数はそれぞれ P, Q とし，範囲は $1 \le p, q \le P, Q$ とする．全極型伝達関数は音声合成で頻繁に用いられるソースフィルタモデルを構成する関数である[6]．

表記の都合上，図 2 中の要素信号を上から $0, 1, 2, \ldots, JK + L - 1$ とナンバリングするインデックス m を導入し，

$$\left(W_{n,f}^m,\, U_n^m\right) := \begin{cases} \left(\dfrac{G_{n,f}^k}{|A_f^j|},\, H_n^{j,k}\right), & \begin{array}{l} k \leftarrow m \bmod K \\ j \leftarrow \lfloor m/K \rfloor \end{array} \quad , \quad (0 \le m < JK) \\ \left(\dfrac{1}{|B_f^l|},\, I_n^l\right), & l \leftarrow m - JK \ , \quad (JK \le m < JK + L) \end{cases} \tag{5}$$

と置く．ただし mod は剰余，$\lfloor \cdot \rfloor$ は床関数である．このとき，式 (2) は次のように書ける．

$$X_{n,f} = \sum_m W_{n,f}^m U_n^m \tag{6}$$

従って $JK + L$ 個のスペクトルパターン $W_{n,f}^m$ と音量 U_n^m によるモデル化と説明できる．ここで，HITF は非負値行列因子分解 (Nonnegative Matrix Factorization: NMF)[7] の拡張と見なせることに注目したい．通常の NMF は，振幅スペクトログラムを，特定の形状に限定しない M 個の (時不変な) スペクトルパターン H_f^m と，時変の音量 U_n^m の積でモデル化する．

$$X_{n,f} := \sum_m H_f^m U_n^m \tag{7}$$

通常の NMF による楽器音分析では調波構造はスペクトルパターンに吸収されるので，基本周波数の異なる音は別個に分解され，また 1 つのパターンに複数音が混合する可能性が高い．あらかじめ調波構造の形状に限定したスペクトルパターンを用いた NMF も報告されているが[4]，異なる音高の単音をまとめることができない問題が残されていた．HITF はこの問題を解消すると同時に類似の包絡構造ごとに分類できるモデル設計になっている．

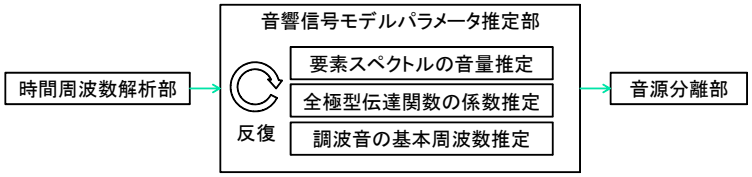

図 3 音源分離処理の概要

2.3 音源分離の手順

音源分離処理は**図 3** のように，まず入力信号とこの数理モデルをもっとも近づけるパラメータ $\{\mu_n^k, \alpha_p^j, \beta_q^l, U_n^m\}$ を推定し[⋆1]，次にその推定結果を用いてユーザ所望の音のみを出力するようなフィルタを作成し適用することで実現する．ひとたび推定結果が得られたら，分離部で U_n^m のうち無音化したい部分を 0 に置き換えた $\tilde{U}_n^m$ による次のようなフィルタリングを行い，出力すべき振幅スペクトログラム $\tilde{Y}_{n,f}$ を得る．

$$\tilde{Y}_{n,f} \leftarrow Y_{n,f} \times \frac{\sum_m W_{n,f}^m \tilde{U}_n^m}{\sum_m W_{n,f}^m U_n^m} \tag{8}$$

この振幅スペクトログラム $\tilde{Y}_{n,f}$ を時間領域信号に戻す処理は，入力信号の位相スペクトログラム $\phi_{n,f}$ を用いて逆 STFT 処理を行う方法と，位相復元法を適用する方法がある[8]．一般的に前者は高速で後者は高品質となる．

音源分離部の実装は平易であり，本手法の実現のための課題の多くは HITF のモデルパラメータをどう推定するかにある．

3. HITF のモデルパラメータ推定

3.1 最適化規準の設定

パラメータ推定は，入力信号と HITF モデルの間の何らかの乖離の度合いを表す関数 Q を最小化する最適化問題として定式化される．

$$\text{minimize } Q\Big(\{Y_{n,f}\}\,,\, \{X_{n,f}\}\Big) \quad \text{w.r.t. } \{\mu_n^k, \alpha_p^j, \beta_q^l, U_n^m\} \tag{9}$$

⋆1 σ^2 は STFT 条件に応じた固定値が使えるが，パラメータ推定時に更新することも可能である[3]．詳細は省略する．

以後 Q のことを最適化規準と呼ぶ．Q は，遂行するタスク (ここでは音源分離) との相性や，パラメータ推定の容易さなどを考慮し具体的に設計することになる．音源分離の問題では以下で定義される I ダイバージェンス $Q^{(\mathrm{I})}$ がよく用いられる[9]．

$$Q^{(\mathrm{I})} := \sum_{n,f} \left(Y_{n,f} \log \frac{Y_{n,f}}{X_{n,f}} - (Y_{n,f} - X_{n,f}) \right) \tag{10}$$

なお，この最適化規準に音量 U_n^m を 0 に近づける (スパース化する) 制約をつけることもできる[5]. I ダイバージェンスは，調波ガウス列の基本周波数パラメータ μ_n^k の更新式が簡潔に導出できることが知られている[2]．一方，全極フィルタ係数 α_p^j, β_q^l は IS ダイバージェンスと呼ばれる別の最適化規準のもとで推定されることが一般的であり[6]，I ダイバージェンスによる推定アルゴリズムは報告されてこなかった．

以下，まず 3.2 節で I ダイバージェンスによる全極型伝達関数の新しいパラメータ推定アルゴリズムを報告する．次に 3.3 節で HITF 全体のパラメータ推定アルゴリズムを示す．

3.2 I ダイバージェンス規準の全極型伝達関数のパラメータ推定

本節では，I ダイバージェンスによる全極型伝達関数のパラメータ推定アルゴリズムを紹介するために，ある時刻の入力振幅スペクトル Y_f を単一の全極型伝達関数で推定するという小課題を考える．したがって，本節中では時刻のインデックス n は省略する．

$$Y_f \simeq \frac{\gamma}{|A_f|} := \frac{\gamma}{|1 - \sum_p \alpha_p \mathrm{e}^{-i\tilde{f}p}|} \tag{11}$$

ここで，γ は本節中でのみ使用する音量パラメータである．最適化規準を I ダイバージェンスとし，まずパラメータ推定に関係しない項を除くと，

$$Q^{(\mathrm{I})} = \sum_f \left(Y_f \log \frac{|A_f|}{\gamma} + \frac{\gamma}{|A_f|} \right) \tag{12}$$

となる．以下，上式から各パラメータの更新式を導出する．

音量パラメータ γ の更新式は，上式の γ の偏微分を 0 と置いて，次のように得られる．

$$\gamma \leftarrow \frac{\sum_f Y_f}{\sum_f |A_f|^{-1}} \tag{13}$$

次に伝達関数の係数 α_p の更新式を導出する．従来の IS ダイバージェンス規準の推定[6] では，最適化規準は α_p に対する二次形式となっていて，その偏微分形から容易に更新式を導出できた．一方 $Q^{(\mathrm{I})}$ は α_p に関する二次形式ではないので，解析的更新は困難である．

そこで，補助関数法[7),10)] を用いて式 (12) を α_p に関する二次形式に変形することを考える．補助関数法とは，最小化したい最適化規準 $Q(\theta)$ に対して次の条件：

$$Q(\theta) = \min_{\vartheta} Q^+(\theta, \vartheta) \tag{14}$$

を満たす補助関数 $Q^+(\theta, \vartheta)$ を設計し，Q^+ に対し補助変数 ϑ に関する最小化と本来の変数 θ に関する最小化を反復することで，間接的に本来の最適化規準を単調減少させる手法である．$Q^+(\theta, \vartheta)$ を最小にする θ, ϑ がともに解析的に解けるように Q^+ を設計すればパラメータ推定は簡単化される．

以下，式 (12) に対する補助関数を設計していく．まず，第 1 項の対数関数による $|A_f|$ の非線形性を解消するために以下の不等式を考える．

$$\frac{1}{2}\log|A_f|^2 \leq \frac{1}{2}\log\rho_f + \frac{1}{2\rho_f}\left(|A_f|^2 - \rho_f\right) = \frac{1}{2\rho_f}|A_f|^2 + \frac{1}{2}\left(\log\rho_f - 1\right) \tag{15}$$

この右辺は凹関数 $\frac{1}{2}\log|A_f|^2$ の点 ρ_f に対する接線であり，ρ_f を補助変数とした補助関数が定義できる．等号成立は $\rho_f \leftarrow |A_f|^2$ としたときであり，これが補助変数の更新式となる．次に，式 (12) 第 2 項の $|A_f|$ の逆数を解消するために第 2 項の点 τ_f の周りの 2 次の Taylor 近似を考える．

$$\frac{1}{|A_f|} \simeq \frac{1}{\tau_f} - \frac{1}{\tau_f^2}\left(|A_f| - \tau_f\right) + \frac{2}{\tau_f^3}\left(|A_f| - \tau_f\right)^2 = \frac{2}{\tau_f^3}|A_f|^2 - \frac{5}{\tau_f^2}|A_f| + \frac{4}{\tau_f} \tag{16}$$

この右辺は必ずしも元の式より大きい値をとるとは限らないので，補助関数の要件を厳密には満たさないが，$\tau_f \leftarrow |A_f|$ と更新すれば凸関数に対する Newton 法と同形になるので，τ_f を補助変数とみた効率的な反復最適化ができる．実際，この補助関数を利用したパラメータ推定は，通常の補助関数と同様に安定して収束することを実験的に確認している．

上の 2 式を用いて，元々の最適化規準に対する補助関数 Q^+ が得られる．

$$\begin{aligned} Q^+ &= \sum_f \left(\frac{Y_f}{2\rho_f}|A_f|^2 + \left(\frac{2}{\tau_f^3}|A_f|^2 - \frac{5}{\tau_f^2}|A_f| \right)\gamma \right) + C \\ &= \sum_f \left(\left(\frac{Y_f}{2\rho_f} + \frac{2\gamma}{\tau_f^3} \right)|A_f|^2 - \frac{5\gamma}{\tau_f^2}|A_f| \right) + C \end{aligned} \tag{17}$$

ただし，C は α_p を含まない項を指す．この時点で式は $|A_f|$ に対して線形になったが，未だ α_p についての 2 次形式とはなっていない．そこでさらに，$|A_f|$ の項に対して複素数の補

Vol.2012-MUS-94 No.27
Vol.2012-SLP-90 No.27
2012/2/4

助変数 ω_f を用いた以下の不等式を考える．

$$-|A_f| \leq -\mathrm{Re}\left[\omega_f^* A_f\right], \qquad |\omega_f| = 1 \tag{18}$$

ここで，$\mathrm{Re}\,[\cdot]$ は実部を示す．これより，更なる補助関数

$$Q^{++} = \sum_f \eta_f \left| A_f - \frac{5\gamma\omega_f}{2\eta_f \tau_f^2} \right|^2 + C = \sum_f \eta_f \left| \frac{\psi_f}{\eta_f} - \sum_p \alpha_p \mathrm{e}^{-i\check{f}p} \right|^2 + C \tag{19}$$

$$\eta_f := \frac{Y_f}{2\rho_f} + \frac{2\gamma}{\tau_f^3}, \qquad \psi_f := \eta_f - \frac{5\gamma\omega_f}{2\tau_f^2} \tag{20}$$

が得られ，α_p についての二次形式に帰着された．

式 (19) を用いた α_p の更新するには，まず 3 つの補助変数を

$$\rho_f \leftarrow |A_f|^2, \qquad \tau_f \leftarrow |A_f|, \qquad \omega_f \leftarrow \frac{A_f}{|A_f|} \tag{21}$$

と更新する．次に式 (19) の α_p による偏微分の実部を 0 と置くことで得られる方程式

$$\sum_{f,q} \eta_f \alpha_q \mathrm{e}^{-i\check{f}(p-q)} = \mathrm{Re}\left[\sum_f \psi_f \mathrm{e}^{-i\check{f}(p)}\right] \tag{22}$$

を，$p = 1, \ldots, P$ まで全て連立して得られる以下の線形方程式によって α_p を更新する．

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_P \end{pmatrix} \leftarrow \begin{pmatrix} R_0 & \cdots & R_{1-P} \\ \vdots & \ddots & \vdots \\ R_{P-1} & \cdots & R_0 \end{pmatrix}^{-1} \begin{pmatrix} r_1 \\ \vdots \\ r_P \end{pmatrix} \tag{23}$$

$$R_p := \sum_f \eta_f \mathrm{e}^{-i\check{f}p}, \qquad r_p := \mathrm{Re}\left[\sum_f \psi_f \mathrm{e}^{-i\check{f}p}\right] \tag{24}$$

となる．これは対称 Toeplitz 型の方程式であり，Levinson-Durbin アルゴリズムを用いて通常の逆行列計算に比べ高速に解くことができる[11]．

3.3 HITF 全体のパラメータ推定アルゴリズム

本節では，HITF 全体のパラメータ推定アルゴリズムについて，前節の全極型伝達関数の係数推定法を参照しつつ述べる．このアルゴリズムは，入力振幅スペクトログラム $Y_{n,f}$ と式 (2, 6) による HITF のモデル $X_{n,f}$ の間の式 (10) の I ダイバージェンス規準最適化である．

3.3.1 音量の更新

まず補助関数を立てる．負の対数関数の中にある和を解消するため以下の Jensen の不等式を考える．

$$-\log \sum_m W_{n,f}^m U_n^m \leq \sum_m \lambda_{n,f}^m \left(-\log \frac{W_{n,f}^m U_n^m}{\lambda_{n,f}^m} \right) \tag{25}$$

ここで，$\lambda_{n,f}^m$ は $\forall n, f, m : \lambda_{n,f}^m > 0$ かつ $\forall n, f : \sum_m \lambda_{n,f}^m = 1$ を満たす変数である．この不等式の等号成立条件は Lagrange の未定乗数法を用いて

$$\lambda_{n,f}^m = \frac{W_{n,f}^m U_n^m}{\sum_m W_{n,f}^m U_n^m} \tag{26}$$

と得られ，これが補助変数の更新式となる．式 (25) により，補助関数

$$Q^+ = \sum_{m,n,f} \left(-Y_{n,f} \lambda_{n,f}^m \log W_{n,f}^m U_n^m + W_{n,f}^m U_n^m \right) + C \tag{27}$$

が立てられる．ただし C はモデルパラメータ $\mu_n^k, \alpha_p^j, \beta_q^l, U_n^m$ を含まない項である．

式 (27) を U_n^m で偏微分した次式

$$\frac{\partial}{\partial U_n^m} Q^+ = -\sum_{m,f} \frac{Y_{n,f} \lambda_{n,f}^m}{U_n^m} + \sum_{m,f} W_{n,f}^m \tag{28}$$

を 0 と置いた方程式から，各要素スペクトルの音量 U_n^m に対する次の更新式が得られる．

$$U_n^m \leftarrow \frac{\sum_{m,f} Y_{n,f} \lambda_{n,f}^m}{\sum_{m,f} W_{n,f}^m} \tag{29}$$

3.3.2 全極型伝達関数のパラメータ更新

今，式 (27) を書き直すと，調波音用の全極型伝達関数のパラメータ α_p^j に関わる部分は

$$\sum_{j,f} \left(\sum_{k,n} Y_{n,f} \lambda_{n,f}^{jK+k} \log |A_f^j| + \frac{\sum_{k,n} G_{n,f}^k H_n^{j,k}}{|A_f^j|} \right) \tag{30}$$

となっており，式 (12) と類似の形状になっている．従って各 j ごとに α_p^j の更新式は，前節の式 (23) による結果に対して $Y_f \leftarrow \sum_{k,n} Y_{n,f} \lambda_{n,f}^{jK+k}$ 及び $\gamma \leftarrow \sum_{k,n} G_{n,f}^k H_n^{j,k}$ を代入したものとなる．一方非調波音用の全極型伝達関数パラメータ β_q^l についても同様である．具体的には $Y_f \leftarrow \sum_n Y_{n,f} \lambda_{n,f}^{JK+l}$，$\gamma \leftarrow \sum_n I_n^l$ として解けばよい．なお α_p^j の更新式を補助

Vol.2012-MUS-94 No.27
Vol.2012-SLP-90 No.27
2012/2/4

変数の更新もまとめて書き下すと以下のような結果となる．

$$\begin{pmatrix} \alpha_1^j \\ \vdots \\ \alpha_P^j \end{pmatrix} \leftarrow \begin{pmatrix} R_0^j & \cdots & R_{1-P}^j \\ \vdots & \ddots & \vdots \\ R_{P-1}^j & \cdots & R_0^j \end{pmatrix}^{-1} \begin{pmatrix} r_1^j \\ \vdots \\ r_P^j \end{pmatrix} \tag{31}$$

$$R_p^j := \sum_f \left[\frac{1}{2|A_f^j|^2} \sum_{k,n} Y_{n,f} \lambda_{n,f}^{jK+k} + \frac{2}{|A_f^j|^3} \sum_{k,n} G_{n,f}^k H_n^{j,k} \right] \mathrm{e}^{-i\check{f}p} \tag{32}$$

$$r_p^j := \mathrm{Re}\left[\sum_f \left[\frac{1}{2|A_f^j|^2} \sum_{k,n} Y_{n,f} \lambda_{n,f}^{jK+k} + \frac{4-5A_f^j}{2|A_f^j|^3} \sum_{k,n} G_{n,f}^k H_n^{j,k} \right] \mathrm{e}^{-i\check{f}p} \right] \tag{33}$$

ただし，上式中の A_f^j は更新前の値を意味するものとする．

3.3.3 基本周波数の更新

基本周波数の更新式は，式 (27) の第 1 項のみを対象として導く．すなわち，第 2 項

$$\sum_{m,n,f} W_{n,f}^m U_n^m \tag{34}$$

は基本周波数に依存しないと仮定する．その理由は次の 2 つである．

(1) この項はガウス関数の値を周波数方向に足し合わせることを表しており，もし周波数毎の加算重みが一定であれば，ガウス関数の平均 (=基本周波数) の位置に関わらず合計値はほぼ一定の値をとる．

(2) 実際には周波数毎の加算重みは一定ではなく全極型伝達関数の形状に応じて変化するが，その形状は滑らかであり，また調波構造用のガウス関数の分散も小さいので，基本周波数の微少な変化ではやはり式 (34) の値の変化は小さい．

今，式 (27) の第 1 項のうち μ_n^k に関わる成分を書き直すと

$$-\sum_{k,j,n,f} Y_{n,f} \lambda_{n,f}^{jK+k} \log \left(\sum_h \exp\left[-\frac{(\hat{f} - h\mu_n^k)^2}{2\sigma^2} \right] \right) \tag{35}$$

である．ここで，Jensen の不等式

$$-\log \left(\sum_h \exp\left[-\frac{(\hat{f} - h\mu_n^k)^2}{2\sigma^2} \right] \right) \leq \sum_h \psi_{n,f}^{h,k} \left(\frac{(\hat{f} - h\mu_n^k)^2}{2\sigma^2} - \log \frac{1}{\psi_{n,f}^{h,k}} \right) \tag{36}$$

表 1 HITF におけるパラメータ推定アルゴリズム．

1. 基本周波数 μ_n^k を 3.4 節の方法で初期化，音量 U_n^m を非負乱数で初期化，全極型伝達関数の係数 α_p^j, β_q^l は $\{Y_{n,f}\}$ 中からランダムに選出した時間フレームのスペクトルに対し，従来法[6] による推定を行い，その結果を初期値とする．
2. 式 (27) 中の $\lambda_{n,f}^m$ を式 (26) で求める．
3. 音量 U_n^m を式 (29) に基づき更新する．
4. 基本周波数 μ_n^k を式 (39) に基づき更新する．
5. 全極型伝達関数の係数 α_p^j, β_q^l を式 (31)〜式 (33) に基づき更新する．
6. 音量 U_n^m の小さくなった調波構造をモデルから除去する．これは計算量の削減につながる．
7. 各パラメータが収束するまで 2 から反復する．

$$\forall h,k,n,f : \psi_{n,f}^{h,k} > 0 \quad \text{and} \quad \forall n,f : \sum_{h,k} \psi_{n,f}^{h,k} = 1 \tag{37}$$

を利用して補助関数

$$Q'^{+} = \sum_{j,k,h,n,f} Y_{n,f} \lambda_{n,f}^{jK+k} \psi_{n,f}^{h,k} \frac{(\hat{f} - h\mu_n^k)^2}{2\sigma^2} + C \tag{38}$$

が得られる．これを μ_n^k で偏微分した式を 0 と置いて，以下の更新式が得られる．

$$\mu_n^k \leftarrow \frac{\sum_{j,h,f} h\hat{f} Y_{n,f} \lambda_{n,f}^{jK+k} \psi_{n,f}^{h,k}}{\sum_{j,h,f} h^2 Y_{n,f} \lambda_{n,f}^{jK+k} \psi_{n,f}^{h,k}} \tag{39}$$

3.4 パラメータ推定の実行

上記アルゴリズムに基づいてパラメータを推定する場合は，基本周波数の初期値依存性の問題を軽減するため，次のような方針をとると良い．まず，調波構造の数 K を想定最大同時発音数よりも大きくとり，基本周波数の値を対数軸で等間隔に並ぶように設定する．次に，パラメータの更新を行い，各時刻ごとに，音量 U_n^m の値が小さくなった調波構造をモデルから除去する．以上のアルゴリズムをまとめると**表 1** のようになる．

4. GUI の実装

提案法を用いて音色ごとの分離結果を任意の比率で再混合できる図 1,4,5 のインターフェースを作成した．一般的なデスクトップミュージックソフトのように，各分離結果はトラック状に可視化され，左側は音色構造を表示し，右側は調波音をピアノロール風の音高-音量形式で，非調波音を音量の時間変化で示している．各トラックはボリュームスライダーとソロ・ミュートボタンを備えている．

Vol.2012-MUS-94 No.27
Vol.2012-SLP-90 No.27
2012/2/4

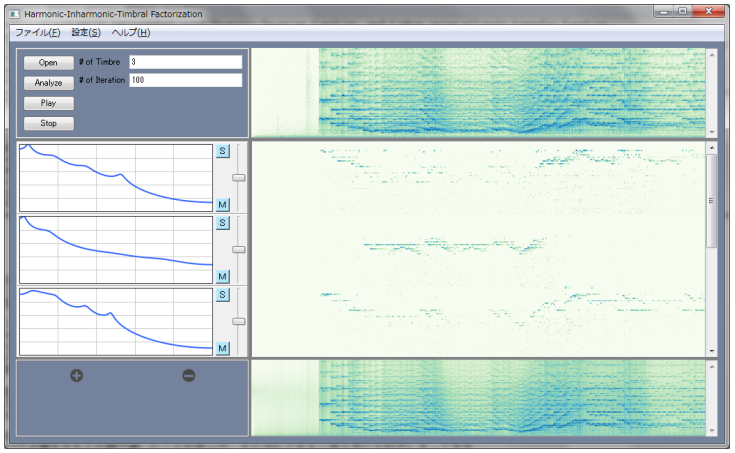

図 4 HITF 動作例 2：バイオリン単旋律を 3 音色で推定した結果.

4.1 動 作 例

■ 1. ピアノとバイオリンの多重奏を 2 音色で推定

本稿冒頭の図 1 がピアノとバイオリンの多重奏を分析した結果であり，ピアノが 1 トラック目，バイオリン部分が 2 トラック目に集中している．したがって，各トラックのミュート/ソロを行うことで楽器パート単位の音源分離が達成される．

■ 2. 単旋律バイオリン演奏を調波 3 音色で推定

図 4 は単旋律のバイオリン演奏を 3 つの全極型伝達関数で分析した結果である．各トラックで際立つ成分が異なり，特に音高ごとにトラックが分かれていることが読み取れる．一般に楽器音のスペクトル包絡構造は音高に若干依存するので，その依存性に合わせて演奏が分離されたと考えられる．

■ 3. ドラムを含むポップス曲を調波 3 音色+非調波 3 音色で推定

図 5 では，非調波成分のトラックに規則的な音量増減が観測できる．ドラムトラックが非調波成分として分離され，また非調波成分 3 トラックの音色形状がバスドラム (低域)，スネア (中域)，ハイハット (高域)，にそれぞれ適応している．ただし，分離結果を視聴する限りでは，調波成分と非調波成分の分離は完全ではないとも感じられた．特にピアノのアタック音はいずれの分離信号からも明確に聞き取ることができた．

4.2 動 作 速 度

本手法および GUI の実装は基本的に Python を用い，HITF のパラメータ推定部分など計算量の多い部分は Cython[12] を用いた．クアッドコア CPU (動作周波数 2.5GHz 程度) の計算機でパラメータ推定を 100 回反復した場合の計算時間は，44.1kHz サンプリングのモノラルデータにおいて，信号長の 6 倍程度であった．計算時間の多くは調波構造を構成する調波ガウス関数列の算出に使われており，この部分の実装に最適化や近似をさらに施すことにより，計算時間をより短縮できると期待される．

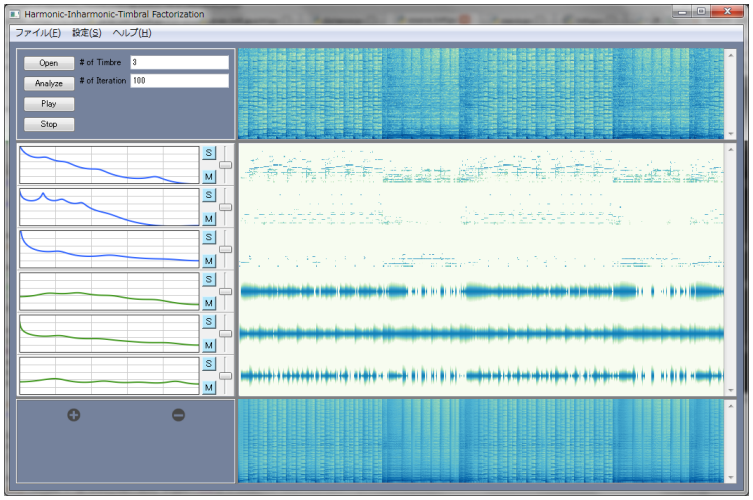

図 5 HITF 動作例 3：ドラムを含む.

5. 評 価 実 験

本手法の音源分離精度に着目した客観的な評価実験について述べる．この実験では，楽器パート数のみを既知として分離を行い，各パート個別の信号と分離結果がどれだけ近いかを評価する．提案法は調波音色数 J を楽器パート数，非調波音色数 L を 0 とした HITF とする．比較法は，通常の NMF を利用した次の手順に基づく自動音源分離法である．

(1) $J \times 10$ 個のスペクトルパターンを持つ NMF で入力音響信号を要素分解する．
(2) 各スペクトルパターンの音色として，従来法[6] により全極スペクトルを得る．
(3) 全極スペクトルを次数 F のベクトルと見て，k-means 法で J 個のクラスに分ける．
(4) クラスタリング結果に基づいて式 (8) と類似の方法で音源分離を行う．すなわち着目したクラスに属する音量パラメータのみ 1, 他を 0 としたモデルのスペクトルを用い

表 2 分離実験結果 SNR (dB): HITF=提案法, NMF=比較法.

楽曲名	楽器パート名	HITF	NMF
Classic #37	Violin	**3.68**	-0.86
	Piano	**12.55**	8.83
Classic #39	Piano	**7.61**	5.90
	Violin	**1.74**	-0.89
Classic #42	Harp	**3.92**	3.00
	Cello	**4.41**	3.66
Jazz #11	Vibraphone	-0.16	**0.11**
	Piano	1.32	**1.38**
Jazz #12	Piano	**7.24**	3.34
	Flute	**3.85**	-0.26
Jazz #14	Piano	**12.56**	4.60
	Bass	**10.41**	8.01

てフィルタリングする．

評価データは RWC Music Database: Jazz Music and Classic Music[13] の両ジャンルから，調波的楽器 2 パートによる演奏を 3 曲ずつ選出した．各曲の Standard MIDI File から MIDI 音源を用いて混合音，各パート個別音をそれぞれ合成し，入力信号及び分離結果の真値とした．主な実験条件については，サンプリング周波数が 44.1kHz, STFT 解析時の窓関数は 2048 点のガウス窓, シフト幅は 512 点とし，全極型伝達関数の次数 P, Q は 10 とした．

分離精度は信号対雑音比 (SNR) を用いる．すなわち，j 番目の楽器パートの真の振幅スペクトログラムを $Y_{j,n,f}$，ξ 番目の分離結果を $X_{\xi,n,f}$ としたとき，

$$\mathrm{SNR}_j := \max_{\xi} \left[10 \log_{10} \frac{\sum_{n,f} Y_{j,n,f}^2}{\sum_{n,f} |Y_{j,n,f} - X_{\xi,n,f}|^2} \right] \tag{40}$$

によって分離の良し悪しを判断する．

表 2 に各分離結果の SNR を示す．Jazz #11 を除いて提案法の方が分離精度が高く，提案法の有効性が示されている．Jazz #11 はピアノとビブラフォンの曲であり，音色の似た減衰音同士の楽器編成だったために，両手法でパート分離に失敗したと考えられる．

6. おわりに

本稿では，多重奏音響信号を調波/非調波音，音高，音色に基づいて分解する新しい音響信号分析法 HITF について報告した．人間の音の聞き分け方に沿ったモデル定義とそのパラメータ推定法を新たに開発し，試作 GUI で操作性を確認後，分離性能の客観評価を行い，通常の NMF を上回る結果を確認した．

今後の課題には，音色の似た楽器編成をより正しく推定できるように，音量・音色の時間変化の情報を取り入れるようにモデルを改良することが挙げられる．通常の NMF に時間変化を導入した手法は既に報告されており[14]，このアプローチとの統合について検討したい．その他更なる課題として，多チャンネル入力時に各要素信号の定位情報を活用する, 音源分離以外のアプリケーション (特に自動採譜) へ展開する，などにも着手したい．

参 考 文 献

1) Smaragdis, P. and Brown, J.: Non-negative matrix factorization for polyphonic music transcription, *Proc. WASPAA*, pp.170–180 (2003).
2) Kameoka, H., Nishimoto, T. and Sagayama, S.: Extraction of multiple fundamental frequencies from polyphonic music using harmonic clustering, *Proc. ICA*, pp.I–59–62 (2004).
3) Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Parameter estimation for harmonic and inharmonic models by using timbre feature distributions, *IPSJ Journal*, Vol.50, No.7, pp.1757–1767 (2009).
4) Vincent, E., Bertin, N. and Badeau, R.: Adaptive harmonic spectral decomposition for multiple pitch estimation, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol. 18, No. 3, pp. 528–537 (2010).
5) Kameoka, H., Ono, N., Kashino, K. and Sagayama, S.: Complex NMF: A new sparse representation for acoustic signals, *Proc. ICASSP*, pp.3437–3440 (2009).
6) Itakura, F. and Saito, S.: Analysis synthesis telephony based on the maximum likelihood method, *Proc. ICA*, pp.C–17–C–20 (1968).
7) Lee, D.D. and Seung, H.S.: Algorithms for non-negative matrix factorization, *Proc. NIPS*, pp.556–562 (2001).
8) Zhu, X., Beauregard, G.T. and Wyse, L.L.: Real-time signal estimation from modified short-time Fourier transform magnitude spectra, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.15, No.5, pp.1645 –1653 (2007).
9) FitzGerald, D., Cranitch, M. and Coyle, E.: On the use of the beta divergence for musical source separation, *Proc. ISSC*, pp.1–6 (2009).
10) Kameoka, H., Ono, N. and Sagayama, S.: Auxiliary function approach to parameter estimation of constrained sinusoidal model for monaural speech separation, *Proc. ICASSP*, pp.29 –32 (2008).
11) Levinson, N.: The Wiener RMS error criterion in filter design and prediction, *Journal of Mathematical Physics*, Vol.25, pp.261–278 (1947).
12) Seljebotn, D.S.: Fast numerical computations with Cython, *Proc. Scipy* (2009).
13) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC music database: popular, classical, and jazz music databases, *Proc. ISMIR*, pp.287–288 (2002).
14) Nakano, M., Roux, J.L., Kameoka, H., Kitano, Y., Ono, N. and Sagayama, S.: Nonnegative matrix factorization with Markov-chained bases for modeling time-varying patterns in music spectrograms, *Proc. LVA/ICA* (2010).