

## Music Factorizer: 音楽音響信号をノート単位で編集できるインタフェース

亀岡 弘 和<sup>†1</sup>      ルルー ジョナトン<sup>†1</sup>  
大石 康 智<sup>†1</sup>      柏野 邦 夫<sup>†1</sup>

本報告では、音楽音響信号をノート単位で編集できるインタフェースの実現を目的とし、音楽音響信号を個々のノートに対応する音響信号に分離する問題に対し、スペクトログラムを階層的なパーツ表現により構成する階層的スパース表現の概念に基づく新しい解法を検討する。音楽のスペクトログラムを、 $I$  個の振幅スペクトルパーツと  $J$  個の時間エンベロープパーツで構成されるものと仮定してスペクトログラムをモデル化し、パラメータの最適化アルゴリズムを導出する。この最適化アルゴリズムでは、各々のスペクトルパーツやエンベロープパーツは、与えられた音響信号を構成する典型的なパーツとなるよう学習される。これらのパーツを個別操作し、その操作を音響信号出力に反映するインタフェース “Music Factorizer” を実装する。

## Music Factorizer: A note-by-note editing interface for music waveforms

HIROKAZU KAMEOKA,<sup>†1</sup> JONATHAN LE ROUX,<sup>†1</sup>  
YASUNORI OHISHI<sup>†1</sup> and KUNIO KASHINO<sup>†1</sup>

The goal of this paper is to develop a note-by-note editing interface for music waveforms. We deal with the problem of decomposing a music signal into signals corresponding to musical notes by introducing a hierarchical sparse representation concept, where music spectrograms are assumed to comprise a limited set of spectrum parts and temporal envelope parts. An iterative algorithm for learning these parts from a given observed spectrogram is derived. The resulting hierarchically structured parts-based representation allows us to manipulate spectrogram components individually corresponding to musical note events.

### 1. はじめに

近年、音楽情報処理分野では、音楽再生中にテンポやリズムを変えたり、特定の楽器音を選択的に音量調節・音質加工したりするなど、音響信号をユーザの要求に合わせて自由にカスタマイズできる音楽鑑賞インタフェースの開発が次々と進められている<sup>1)–6)</sup>。本研究では、この一連の研究の取り組みの一環として、音楽音響信号をあたかも MIDI シーケンスソフトで MIDI データを扱うかのような感覚で自由に編集できるインタフェースの実現を目指しており、本報告ではそれに向けた基礎検討を行う。

ミックスダウンされた音楽音響信号に対し、各ノートに対応する音響信号ごとの操作や加工を可能にするためには、個々の信号に正確に分解する必要があるが、一般の混合信号から個々の信号を分離抽出することは容易ではない。このような問題は音源分離と呼ばれ、特にモノラル信号を対象とした音源分離は典型的な不良設定問題であり、何の仮定も置かずには解くことは困難である。このため、いかに一般性のある仮定を置いてうまく解を絞り込めるかが重要となる。モノラル音源分離の研究は、これまで多くの研究者によって検討されているが、そのアプローチは、問題設定の際に立てられる仮定に応じて大別できる。例えば代表的なアプローチは、音の調波性、周波数成分の時間方向の連続性など、時間周波数成分の各時刻における局所的性質を各音源に対して仮定した手法である。時間周波数成分の局所的性質は、人間が音を聴き分ける際の有用な手がかりとなっていると考えられていることから、このタイプのアプローチは人間の聴覚機能を模倣する試みとして計算論的聴覚情景分析と呼ばれている。一方、近年有効なアプローチとして脚光を浴びているのは、スパース表現と呼ぶ方法論を応用した手法である。このアプローチは、音源に関する仮定をほとんど置かない代わりに、音響信号の全体が限られた種類の要素だけで構成されていると見なし、その要素を情報論的な規準に基づいて推定しようというものである。本報告では、後者のスパース表現の考え方をヒントにし、音楽音響信号に適した新しい信号表現方法を提唱する。そしてこの信号表現方法に基づき、音楽音響信号の新しいモデルを提案し、さらにその最適化アルゴリズムを導く。この手法をもとにして、モノラル音楽音響信号をノート単位で編集できる新しいタイプの音楽鑑賞インタフェース “Music Factorizer” を実現する。

<sup>†1</sup> 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 Communication Science Laboratories, Nippon Telegraph and Telephone Corporation

## 2. 音楽の階層的スパース表現

音楽は、多数のメロディーの音響信号が多層に重畳した複雑な音響信号を形成するが、一方で独特の規則性をもっている。例えば、多くの楽曲においては、音階に対応する限られた種類の基本周波数の音だけで構成される、限られた種類の楽器の音だけで構成される、限られた種類の音長（八分音符、四分音符など）の音だけで構成される、限られた種類のメロディーやフレーズが時間を隔てて繰り返される、限られた種類のセクション（「A メロ」、「サビ」など）が繰り返される、といった特徴がある。すなわち、音楽音響信号は、さまざまな階層において限られた種類の「パーツ」で構成されている、という性質をもっており、この大域的な性質を拘束として考えれば、局所的な観測だけでは難しい不良設定な音源分離の問題も、全体の観測を通して効果的に解ける可能性がある。この観点に立って、以下で音楽音響信号のモデル化を行う。

### 2.1 スペクトログラムモデル

本報告で提案するモデルは、非負値行列因子分解 (NMF)<sup>7)</sup> によるスペクトログラムの分解表現<sup>8)</sup> を拡張したものである。NMF によるスペクトログラムの分解表現は、スペクトログラムを非負値行列  $\mathbf{Y} = (Y_{\omega,t})_{\Omega \times T} \in \mathbb{R}^{\geq 0, \Omega \times T}$  と見立て、

$$\mathbf{Y} \simeq \mathbf{H}\mathbf{U} \quad (1)$$

となるような非負値行列  $\mathbf{H} = (H_{\omega,i})_{\Omega \times I} \in \mathbb{R}^{\geq 0, \Omega \times I}$ ,  $\mathbf{U} = (U_{i,t})_{I \times T} \in \mathbb{R}^{\geq 0, I \times T}$  を決めてやることによって得られる。ただし、 $\omega, t$  はそれぞれ周波数と時刻に対応するインデックスである。 $\mathbf{y}_t := (Y_{1,t}, \dots, Y_{\Omega,t})^T$ ,  $\mathbf{h}_i := (H_{1,i}, \dots, H_{\Omega,i})^T$  とすると式 (1) は、

$$\forall t, \mathbf{y}_t \simeq \sum_{i=1}^I U_{i,t} \mathbf{h}_i, \forall i, U_{i,t} \geq 0, \mathbf{h}_i \geq 0 \quad (2)$$

と書き直せることから分かる通り、すべての  $t$  における観測データ  $\mathbf{y}_t$  を、高々  $I$  種類の「パーツ」 $\mathbf{h}_1, \dots, \mathbf{h}_I$  だけで構成されたものと見なそうとしていることになり、そのためにそれぞれのパーツをどのように置くのが最も妥当かを判断する問題と理解される。これにより得られるそれぞれの行列が表すものは、図 1 を見るとより分かりやすい。 $\mathbf{H}$  の各列ベクトルには楽曲中に繰り返し現れるスペクトルが典型的な構成パーツと見なされて表出される。従って、音楽スペクトログラムが楽器の種類や音階で決まる限られたパターンのスペクトルだけで構成されているとするなら、 $\mathbf{H}$  の各列ベクトルはおおよそ特定楽器の特定音階に対応したスペクトルとなる。一方、 $\mathbf{U}$  の各行ベクトルはそれぞれのスペクトルパーツが

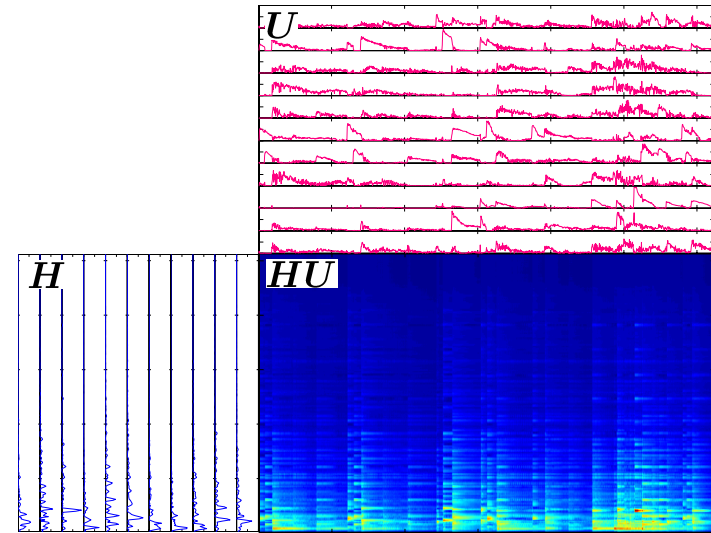


図 1 非負値行列因子分解 (NMF) による音楽スペクトログラムの分解表現

どの時刻にどの程度の強さでアクティベートしているかを表している。

図 1 の  $\mathbf{U}$  の各行ベクトルを見ると、 $\mathbf{h}_i$  のアクティベーションの時間エンベロープは、限られた種類の局所パターンだけで構成されていることに気づく。これは、個々のノートは限られた種類の音長で弾かれるため、立ち上がり・減衰パターンの種類もまた限られるからである。そこで、 $\mathbf{U}$  に関して、さらに、

$$U_{i,t} = \sum_{j=1}^J \sum_{\tau=0}^M G_{j,\tau} O_{i,j,t-\tau}, G_{j,\tau} \geq 0, O_{i,j,t} \geq 0 \quad (3)$$

のような畳み込み混合の形で表される分解表現を考えよう。 $G_{j,\tau}$  は  $j$  番目の時間エンベロープの局所パターンを表す。一方  $O_{i,j,t}$  はそのアクティベーションを表し、理想的には各ノートのオンセット時刻にピークが立つようなイメージとなる。式 (1) の右辺に式 (3) を代入すると、

$$X_{\omega,t} = \sum_i H_{\omega,i} \sum_j \sum_{\tau} G_{j,\tau} O_{i,j,t-\tau} \quad (4)$$

のように、式 (1) で与えられる NMF 型のモデルを拡張したスペクトログラムモデルが立て

られる。ここで、分解のスケールの任意性を除く目的で

$$\sum_{\omega} H_{\omega,i} = 1 \quad (i = 1, \dots, I), \quad \sum_{\tau} G_{j,\tau} = 1 \quad (j = 1, \dots, J) \quad (5)$$

を仮定しておく。以上では、音楽のスペクトログラムは  $I$  個の振幅スペクトルパーツと  $J$  個の時間エンベロープパーツで構成されるという仮定に基づいてスペクトログラムをモデル化した。以上のようにしてスペクトログラムは階層的なパーツ表現により構成していくことができ、今回のモデルはその一例である。このような形態のモデルに基づく分解表現を階層的スパース表現と命名する。今後はメロディー・フレーズ・セクションなど各階層でのパーツ表現を取り入れた、より大域的なモデル化の検討を進めていく予定である。

## 2.2 最適化アルゴリズム

さて、NMF で  $Y$  から図 1 のように  $H$  と  $U$  を求めたように、提案するモデルのもとで  $Y$  から  $H$  と  $G$  と  $O$  を求めたい。以下では、提案モデルによる音楽の階層的スパース表現を得るための最適化アルゴリズムについて述べる。

### 2.2.1 二乗誤差規準

ここでは、観測スペクトログラム  $Y := (Y_{\omega,t})_{\Omega \times T}$  のもとで設定される次の最適化問題

$$\begin{aligned} \text{minimize} \quad & \mathcal{F}(H, G, O) := \sum_{\omega,t} \left| Y_{\omega,t} - \sum_i H_{\omega,i} \sum_j \sum_{\tau} G_{j,\tau} O_{i,j,t-\tau} \right|^2 + S(G, O) \\ \text{subject to} \quad & \forall_{\omega,i} H_{\omega,i} \geq 0, \quad \forall_{j,\tau} G_{j,\tau} \geq 0, \quad \forall_{i,j,t} O_{i,j,t} \geq 0, \\ & \forall_i \sum_{\omega} H_{\omega,i} = 1, \quad \forall_j \sum_{\tau} G_{j,\tau} = 1 \end{aligned}$$

を検討する。 $S(G, O)$  は  $G, O$  をスパースな解へと誘導する正則化項であり、ここでは

$$S(G, O) := 2\lambda_g \sum_j \sum_{\tau} |G_{j,\tau}|^{p_g} + 2\lambda_o \sum_i \sum_j \sum_t |O_{i,j,t}|^{p_o} \quad (6)$$

と定義する。ただし、 $0 < p_g \leq 2, 0 < p_o \leq 2$  とする。以下、最適化アルゴリズムの導出には、NMF<sup>7)</sup> や調波時間構造化クラスタリング<sup>9),10)</sup> などと同様、補助関数法を用いる。

まず、 $\mathcal{F}$  を降下させる  $H$  の更新式を導出する。 $H, G, O$  の 1 ステップ前での更新値をそれぞれ  $H', G', O'$  としよう。すると、

$$\mathcal{F}(H, G', O') \leq \mathcal{F}_H(H, G', O') = \sum_{\omega,t} \sum_i \frac{H'_{\omega,i} U'_{i,t}}{X'_{\omega,t}} \left| Y_{\omega,t} - \frac{H_{\omega,i}}{H'_{\omega,i}} X'_{\omega,t} \right|^2 + S(G, O) \quad (7)$$

が言える。ただし、

$$U'_{i,t} = \sum_j \sum_{\tau} G'_{j,\tau} O'_{i,j,t-\tau} \quad (8)$$

$$X'_{\omega,t} = \sum_i H'_{\omega,i} \sum_{\tau} G'_{j,\tau} O'_{i,j,t-\tau} \quad (9)$$

である。詳細は省略するが、 $\mathcal{F}_H(H, G', O')$  を最小化する  $H$  は解析的に

$$H_{\omega,i} = H'_{\omega,i} \frac{\sum_t U'_{i,t} Y_{\omega,t}}{\sum_t U'_{i,t} \sum_k H'_{\omega,k} U'_{k,t}} \quad (10)$$

と求まり、このように更新すれば  $\mathcal{F}(H, G', O')$  は増加しないことが保証される。また、 $H',$  および  $U'$  がともに非負値であれば  $H$  も必ず非負値となる。

次に、 $\mathcal{F}$  を降下させる  $G$  の更新式を導出する。先と同様、

$$\begin{aligned} \mathcal{F}(H', G, O') \leq \mathcal{F}_G(H', G, O') &= \sum_{\omega,t} \sum_j \sum_{\tau} \frac{G'_{j,\tau} W'_{j,\omega,t-\tau}}{X'_{\omega,t}} \left| Y_{\omega,t} - \frac{G_{j,\tau}}{G'_{j,\tau}} X'_{\omega,t} \right|^2 \\ &+ \lambda_g \sum_j \sum_{\tau} \left[ p_g |G'_{j,\tau}|^{p_g-2} G_{j,\tau}^2 + (2-p_g) |G'_{j,\tau}|^{p_g} \right] + 2\lambda_o \sum_i \sum_j \sum_t |O_{i,j,t}|^{p_o} \end{aligned} \quad (11)$$

が言える。ただし、

$$W'_{j,\omega,t-\tau} = \sum_i H'_{\omega,i} O'_{i,j,t-\tau} \quad (12)$$

である。 $\mathcal{F}_G(H', G, O')$  の第 2 項は

$$|x|^p \leq \frac{p|x'|^{p-2}}{2} x^2 + \left(1 - \frac{p}{2}\right) |x'|^p, \quad (0 < p \leq 2) \quad (13)$$

の不等式に由る (右辺は接点  $\pm x'$  で  $|x|^p$  に接する放物線であるため明らか)。以上より、 $\mathcal{F}_G(H', G, O')$  を用いて  $G$  の更新式が

$$G_{j,\tau} = G'_{j,\tau} \frac{\sum_i \sum_t O'_{i,j,t-\tau} \sum_{\omega} H'_{\omega,i} Y_{\omega,t}}{\sum_i \sum_t O'_{i,j,t-\tau} \sum_{\omega} H'_{\omega,i} X'_{\omega,t} + \lambda_g p_g |G'_{j,\tau}|^{p_g-1}} \quad (14)$$

と導ける。最後に、 $\mathcal{F}$  を降下させる  $O$  の更新式を導出する。先と同様、

$$\begin{aligned} \mathcal{F}(H', G', O) \leq \mathcal{F}_O(H', G', O) &= \sum_{\omega, t} \sum_{i, j} \sum_{\tau} \frac{H'_{\omega, i} G'_{j, t-\tau} O'_{i, j, \tau}}{X'_{\omega, t}} \left| Y_{\omega, t} - \frac{O'_{i, j, \tau}}{O'_{i, j, \tau}} X'_{\omega, t} \right|^2 \\ &+ 2\lambda_g \sum_j \sum_{\tau} |G'_{j, \tau}|^{p_g} + \lambda_o \sum_{i, j} \sum_t \left[ p_o |O'_{i, j, t}|^{p_o-2} O_{i, j, t}^2 + (2 - p_o) |O'_{i, j, t}|^{p_o} \right] \end{aligned} \quad (15)$$

が言え、これを用いて  $O$  の更新式が

$$O_{i, j, \tau} = O'_{i, j, \tau} \frac{\sum_t G'_{j, t-\tau} \sum_{\omega} H'_{\omega, i} Y_{\omega, t}}{\sum_t G'_{j, t-\tau} \sum_{\omega} H'_{\omega, i} X'_{\omega, t} + \lambda_o p_o |O'_{i, j, \tau}|^{p_o-1}} \quad (16)$$

と導ける。

### 2.2.2 I ダイバージェンス規準

モデル化誤差を  $I$  ダイバージェンス<sup>11)</sup> で測った場合の最適化問題

$$\begin{aligned} \text{minimize } \mathcal{I}(H, G, O) &:= \sum_{\omega, t} \left[ Y_{\omega, t} \log \frac{Y_{\omega, t}}{X_{\omega, t}} - (Y_{\omega, t} - X_{\omega, t}) \right] + \mathcal{S}(G, O) \\ \text{where } X_{\omega, t} &:= \sum_i H_{\omega, i} \sum_j \sum_{\tau} G_{j, \tau} O_{i, j, t-\tau} \\ \text{subject to } \forall_{\omega, i} H_{\omega, i} &\geq 0, \forall_{j, \tau} G_{j, \tau} \geq 0, \forall_{i, j, t} O_{i, j, t} \geq 0, \\ \forall_i \sum_{\omega} H_{\omega, i} &= 1, \forall_j \sum_{\tau} G_{j, \tau} = 1 \end{aligned}$$

についても検討する。

2.2.1 と同様、補助関数法を適用する。まず、 $H$  の更新式については、

$$\begin{aligned} \mathcal{I}(H, G', O') \leq \mathcal{I}_H^+(H, G', O') &:= \sum_{\omega, t} \left[ Y_{\omega, t} \log Y_{\omega, t} - Y_{\omega, t} + \sum_i H_{\omega, i} U'_{i, t} \right. \\ &\left. - Y_{\omega, t} \sum_i \frac{H'_{\omega, i} U'_{i, t}}{X'_{\omega, t}} \log \frac{H_{\omega, i}}{H'_{\omega, i}} X'_{\omega, t} \right] + \mathcal{S}(G, O) \end{aligned} \quad (17)$$

の不等式より導かれる補助関数  $\mathcal{I}_H^+$  を最小化する  $H$  を求めれば良く、

$$H_{\omega, i} = H'_{\omega, i} \frac{\sum_t Y_{\omega, t} U'_{i, t} / X'_{\omega, t}}{\sum_t U'_{i, t}} \quad (18)$$

のように解析的に得られる。以下同様にして、 $G$  の更新式については、

$$\begin{aligned} \mathcal{I}(H', G, O') \leq \mathcal{I}_G^+(H', G, O') &:= \sum_{\omega, t} \left[ Y_{\omega, t} \log Y_{\omega, t} - Y_{\omega, t} \right. \\ &\left. + \sum_{j, \tau} G_{j, \tau} W'_{i, j, t-\tau} - Y_{\omega, t} \sum_{j, \tau} \frac{G'_{j, \tau} W'_{j, \omega, t-\tau}}{X'_{\omega, t}} \log \frac{G_{j, \tau}}{G'_{j, \tau}} X'_{\omega, t} \right] \\ &+ 2\lambda_g \sum_{j, \tau} [p_g |G'_{j, \tau}|^{p_g-1} (G_{j, \tau} - G'_{j, \tau}) + |G'_{j, \tau}|^{p_g}] + 2\lambda_o \sum_{i, j, t} |O'_{i, j, t}|^{p_o} \end{aligned} \quad (19)$$

の不等式、 $O$  の更新式については、

$$\begin{aligned} \mathcal{I}(H', G', O) \leq \mathcal{I}_O^+(H', G', O) &:= \sum_{\omega, t} \left[ Y_{\omega, t} \log Y_{\omega, t} - Y_{\omega, t} \right. \\ &\left. + \sum_{i, j, \tau} H'_{\omega, i} G'_{j, t-\tau} O_{i, j, \tau} - Y_{\omega, t} \sum_{i, j, \tau} \frac{H'_{\omega, i} G'_{j, t-\tau} O'_{i, j, \tau}}{X'_{\omega, t}} \log \frac{O_{i, j, \tau}}{O'_{i, j, \tau}} X'_{\omega, t} \right] \\ &+ 2\lambda_g \sum_{j, \tau} |G'_{j, \tau}|^{p_g} + 2\lambda_o \sum_{i, j, t} [p_o |O'_{i, j, t}|^{p_o-1} (O_{i, j, t} - O'_{i, j, t}) + |O'_{i, j, t}|^{p_o}] \end{aligned} \quad (20)$$

の不等式より導かれる補助関数  $\mathcal{I}_G^+$ ,  $\mathcal{I}_O^+$  をそれぞれ最小化する  $G, O$  を求めれば良く、

$$G_{j, \tau} = G'_{j, \tau} \frac{\sum_i \sum_t O'_{i, j, t-\tau} \sum_{\omega} H'_{\omega, i} Y_{\omega, t} / X'_{\omega, t}}{\sum_i \sum_t O'_{i, j, t-\tau} + 2\lambda_g p_g |G'_{j, \tau}|^{p_g-1}} \quad (21)$$

$$O_{i, j, \tau} = O'_{i, j, \tau} \frac{\sum_t G'_{j, t-\tau} \sum_{\omega} H'_{\omega, i} Y_{\omega, t} / X'_{\omega, t}}{\sum_t G'_{j, t-\tau} + 2\lambda_o p_o |O'_{i, j, \tau}|^{p_o-1}} \quad (22)$$

のように解析的に得られる。式 (19) の第 2 項および式 (20) の第 3 項は

$$\forall_{x>0}, |x|^p \leq p|x'|^{p-1}(x-x') + |x'|^p \quad (0 < p \leq 2) \quad (23)$$

に由る (右辺は接点  $x'$  における  $|x|^p$  の接線であるため明らか)。

### 3. Music Factorizer の実装

以上の手法により  $H, G, O$  がひとたび求めれば、スペクトログラム  $X$  を式 (9) に従って  $H, G, O$  の操作を介して加工することができる。提案法を核として、音楽音響信号をノート

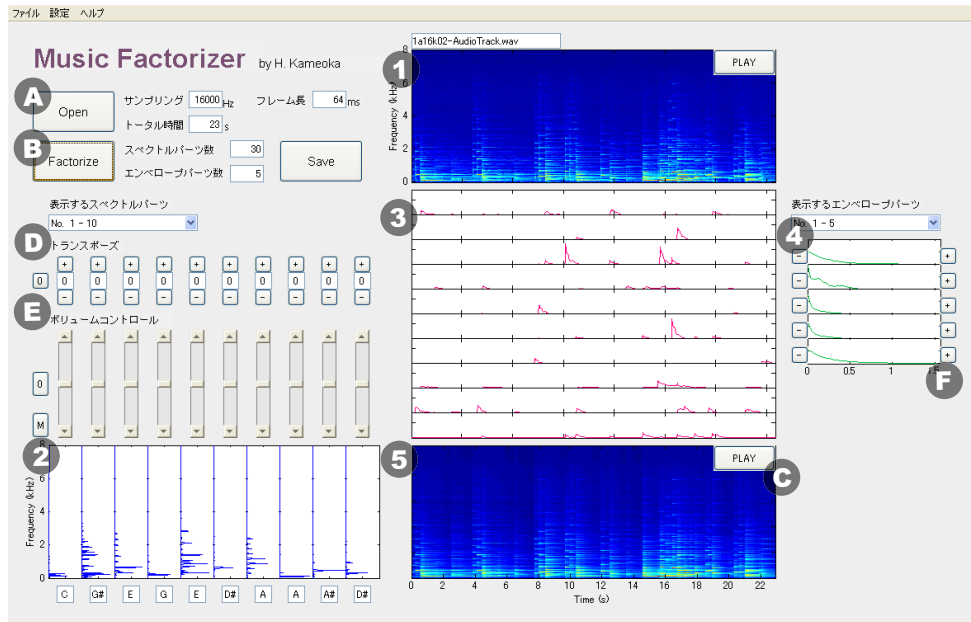


図 2 Music Factorizer の表示画面。①～⑤ のウィンドウと ①～⑥ の操作方法に関する説明は 3 章参照。

単位で編集できるインタフェース “Music Factorizer” を実装した。図 2 にグラフィカルユーザインタフェース (GUI) の表示画面を示す。操作方法および機能は以下のとおりである。

**A: 音響信号の読み込みと時間周波数解析** 音響信号を読み込み、短時間フーリエ変換 (STFT) により時間周波数解析を行ってスペクトログラム  $\mathbf{Y} = (Y_{\omega,t})_{\Omega \times T}$  を出力する。出力された  $\mathbf{Y}$  は図 2 におけるウィンドウ①上にカラー表示される。STFT ではフレーム長 64ms の Hanning 窓をシフト幅 32ms により分析した。

**B: 提案モデルのパラメータ最適化アルゴリズムの実行** 2 章で導入したスペクトログラムモデルとそのパラメータを推定する反復アルゴリズムを実行する。処理の流れは以下のとおりである。

(1) 従来型 NMF<sup>7)</sup> により  $\mathbf{Y}$  から  $\mathbf{H}_{\text{NMF}}$ ,  $\mathbf{U}_{\text{NMF}}$  を出力する。

(2) 提案モデルにおけるパラメータ  $H$ ,  $G$ ,  $O$  の初期値  $H^{(0)}$ ,  $G^{(0)}$ ,  $O^{(0)}$  をそれぞれ

$$\forall \omega, i, H_{\omega,i}^{(0)} = [\mathbf{H}_{\text{NMF}}]_{\omega,i} \quad (24)$$

$$\forall j, \tau, G_{j,\tau}^{(0)} = \exp(-\gamma_j \tau) / \sum_{\tau'} \exp(-\gamma_j \tau') \quad (\gamma_j \text{ は正の乱数}) \quad (25)$$

$$\forall i, j, t, O_{i,j,t}^{(0)} = [\mathbf{U}_{\text{NMF}}]_{i,t} \quad (26)$$

と置く。また、 $U^{(0)}$  および  $X^{(0)}$  は式 (8), (9) により算出する。

(3) 式 (10), 式 (14), 式 (16) により  $H$ ,  $G$ ,  $O$  を反復的に更新していき、反復計算の回数が 50 に到達したらその時点での  $H$ ,  $G$ ,  $O$  の値を出力する。なお、式 (10), 式 (14), 式 (16) の代わりに式 (18), 式 (21), 式 (22) により更新を行うモードに切り替えることもできる。

(4) 出力された  $H$  を図 2 におけるウィンドウ②に、出力された  $G$  と  $O$  を用いて式 (8) により算出される  $U$  をウィンドウ③に、出力された  $G$  をウィンドウ④にそれぞれ表示する。また、式 (9) により算出される  $X$  をウィンドウ⑤にカラー表示する。

図 2 には、提案法の実際の実行結果の例を示しているが、①と⑤を見比べてみると、観測スペクトログラム  $\mathbf{Y}$  が最適化アルゴリズムを通してスペクトログラムモデル  $\mathbf{X}$  によって適切に再現されていることが確認できる。

**C: 音響信号の生成** スペクトログラム  $\mathbf{X}$  から音響信号を Griffin-Lim の反復 STFT 法<sup>13)</sup> に基づいて合成し、再生する。Griffin-Lim の反復 STFT 法とは以下のステップを反復的に行うものである。

- (0)  $H$ ,  $G$ ,  $O$  から構成されるスペクトログラム  $X_{\omega,t} \in \mathbb{R}^{\geq 0}$  に対応する位相をランダムに設定し、位相が付加された複素数値の時間周波数成分を  $V_{\omega,t} \in \mathbb{C}$  とする。
- (1)  $\{V_{\omega,t}\}_{\omega \in \{1, \dots, \Omega\}, t \in \{1, \dots, T\}}$  に対し逆 STFT を行い、実信号  $\{v[n]\}_{n=1}^N$  を得る。
- (2) 実信号  $\{v[n]\}_{n=1}^N$  に対し STFT を行い、 $\{V_{\omega,t}\}_{\omega \in \{1, \dots, \Omega\}, t \in \{1, \dots, T\}}$  を更新する。
- (3) すべての  $\omega, t$  について  $V_{\omega,t} \leftarrow X_{\omega,t} \frac{V_{\omega,t}}{|V_{\omega,t}|}$  により  $V_{\omega,t}$  を更新し、(1) に戻る。

なお、上記のやり方以外に Le Roux の手法<sup>14)</sup> を用いることもできる。

**D: トランスポーズ機能** ②に表示される個々のスペクトルパーツを +/- ボタンにより 100cent(半音) 単位でトランスポーズできる。ここで言うトランスポーズとはスペクトルを周波数方向に線形に伸縮する操作をさす。②の下部には、各スペクトルパーツがど

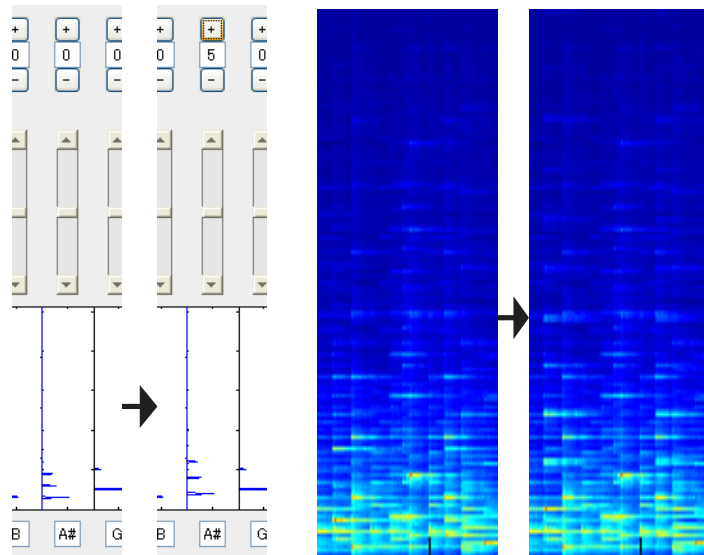


図 3 トランスポーズ機能を用いたスペクトログラム加工例。特定のスペクトル部分を 5 半音分だけトランスポーズした結果、対応する一部の時間周波数成分だけを操作できていることが確認できる。

の音高に最も対応しているらしいかを Subharmonic Summation (SHS)<sup>12)</sup> により推定した結果を表示している。この音高表示を手がかりにすれば、例えば長調の曲に信号を加工するためにはどのスペクトル部分をどのようにトランスポーズすれば良いかといったようなことが判断できるようになる。図 3 に、このトランスポーズ機能を用いたスペクトログラム加工の例を示す。特定のスペクトル部分を 5 半音分だけトランスポーズした結果、対応する一部の時間周波数成分だけを操作できていることが確認できる。

**E: ボリュームコントロール機能** 各スペクトルパーツに対応するノートの音量をスライダーにより自由に調節できる。

**F: サステイン・リリースタイムコントロール機能** 各エンベロープパーツに対応するノートのサステインないしリリースタイムを +/- ボタンにより自由に調節できる。またこの他に、推定された  $O$  に対し、オンセット時刻を表すピーク成分の位置を操作することで各ノートの発音開始のタイミングをコントロールすることも可能である。

#### 4. おわりに

本報告では、音楽音響信号をノート単位で編集できるインターフェースの実現を目的とし、音楽音響信号を個々のノートに対応する信号に分離する問題に対し、スペクトログラムを階層的なパーツ表現により構成する階層的スパース表現の概念に基づく新しい解法を検討した。今後は、インターフェースの機能を充実化を図るとともに、振幅スペクトル領域で立てた本提案モデルを、複素 NMF<sup>15)</sup> と同様、複素スペクトル領域のモデルに拡張していきたい。

#### 参 考 文 献

- 1) K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H.G. Okuno, "Drumix: an audio player with real-time drum-part rearrangement functions for active music listening," *IPSJ Journal* (情報処理学会論文誌), Vol.48, No.3, pp.1229-1239, 2007.
- 2) J. Woodruff, B. Pardo, and R. Dannenberg, "Remixing stereo music with score-informed source separation," *Proc. ISMIR'06*, pp.314-319, 2006.
- 3) 亀岡, 後藤, 嵯峨山, "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, pp.77-84, 2006.
- 4) 糸山, 後藤, 駒谷, 尾形, 奥乃, "楽譜情報を援用した多重奏音楽音響信号の音源分離と調波・非調波統合モデルの制約付パラメータ推定の同時実現," 情報処理学会論文誌, Vol.49, No.3, pp.1465-1479, 2008.
- 5) N. Ono, K. Miyamoto, H. Kameoka, S. Sagayama, "A real-time equalizer of harmonic and percussive components in music signals," *Proc. ISMIR'08*, 139-144, 2008.
- 6) 水野, ルルー, 小野, 嵯峨山, "パワースペクトログラムの伸縮と無矛盾相付加に基づく音楽音響信号の実時間テンポ/ピッチ変換," 日本音響学会春季研究発表会講演論文集, 2-8-4, pp.843-844, 2009.
- 7) D.D. Lee, and H.S. Seung, "Algorithms for nonnegative matrix factorization," *Proc. NIPS'00*, pp.556-562, 2000.
- 8) P.Smaragdis, and J.C. Brown, "Non-negative matrix factorization for music transcription," *Proc. WASPAA'03*, 177-180, 2003.
- 9) H. Kameoka, T. Nishimoto, S. Sagayama, "A multipitch analyzer based on harmonic temporal structured clustering," *IEEE Trans. ASLP*, **15**(3), 982-994, 2007.
- 10) J. Le Roux, H. Kameoka, N. Ono, A. de Cheveigne, S. Sagayama, "Single and multiple F0 contour estimation through parametric spectrogram modeling of speech in noisy environments," *IEEE Trans. ASLP*, **15**(4), pp.1135-1145, 2007.
- 11) I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *The Annals of Probability*, Vol.3, No.1, pp.146-158, 1975.
- 12) D.J. Hermes, "Measurement of pitch by subharmonic summation," *Journal of Acoustic of Society of America*, Vol.83, pp.257-264, 1988.
- 13) D.W. Griffin, and J.S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. ASSP*, **32**(2), pp.236-243, 1984.
- 14) J. Le Roux, N. Ono, S. Sagayama, "Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction," *Proc. SAPA'08*, 2008.
- 15) 亀岡, 小野, 柏野, 嵯峨山, "複素 NMF: 新しいスパース信号分解表現と基底系学習アルゴリズム," 日本音響学会秋季研究発表会講演論文集, 2-8-13, pp.657-660, 2008.