

調波パラメトリック NMF による楽器演奏音響信号の分析合成

安良岡 直希 糸山 克寿 高橋 徹 尾形 哲也 奥乃 博

京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

本稿では調波パラメトリック Nonnegative Matrix Factorization (HPNMF) と呼ぶ新しい振幅スペクトログラム表現法を用いた演奏分析合成法について述べる。我々の最終目的は音色・演奏表情を制御しうる楽器演奏合成系を実現することである。楽器演奏合成とは、楽譜を入力としその演奏音響信号を出力することである。代表的な楽器演奏合成装置である MIDI 音源は、一般に予め録音した楽器単音を再生する方式であり、人間の演奏中の音量や音色の揺らぎは原則含まれない。そこで、特定の実演奏音響信号を学習データに用いてその演奏の特徴を備えた楽器演奏合成系を構築できる枠組みが求められる。

楽器演奏合成系を学習データから獲得する方法として素片選択アプローチと回帰アプローチの2つが考えられる。前者では、演奏音響信号から抽出した各単音（あるいはその音響特徴）と楽譜の対応関係を蓄積し、合成時には楽譜の各単音に対して適切な素片を選択する [1]。しかし素片の選択基準の策定は困難である。一方後者では楽譜を各時間フレーム毎にベクトルデータ（以下楽譜情報ベクトルと呼ぶ）に変換し、これを説明変数とみて音響特徴を算出する。しかし、楽器演奏音に対する効果的な音響特徴と回帰式の設計方法は明らかにされていない。

HPNMF は調波楽器音の新しい振幅スペクトログラム表現法であり、回帰アプローチに基づく楽器演奏合成系の構築に利用できる。HPNMF は、調波構造を表す関数形により各時刻の振幅スペクトルをモデル化した上でその各倍音強度を演奏全体で因子分解する。これにより、楽器音を基本周波数 (F0) と数種の調波成分・非調波成分パターンという音色・演奏表情に大きく関係するパラメータで表せる。楽譜情報ベクトルからの回帰式で HPNMF を用いれば、楽譜から上記パラメータを経て直接振幅スペクトルを算出する効果的な楽器演奏合成系となる。

2. 問題定義と回帰アプローチの概要

本稿で扱う問題は、学習データとして与えられる楽譜と実演奏音響信号の組を用いて、任意の楽譜に対する楽器演奏音響信号を合成することである。合成結果は学習データが持つ音色や演奏表情を再現し、音響的に実演奏のようだと考えることが求められる。現時点では調波楽器の単旋律演奏のみを対象とし、また楽譜と音響信号は時間の同期が取れているとする。

我々は「楽譜から振幅スペクトルを直接算出する」回帰アプローチのもとで上記問題に対処する。何らかの音響特徴ではなく振幅スペクトルを直接回帰する動機は、1) 楽器演奏中の主要な音響特徴 (F0, 倍音強度等) の推定自体が容易ではないため、2) 将来の展望として、多重奏や残響付加済のデータに対して目的演奏を分離しながらの学習を可能にするため、などである。

HPNMF の説明に入る前に、上記方針に基づく回帰モデルの基本形として NMF [2] を用いた実装を示す。通常

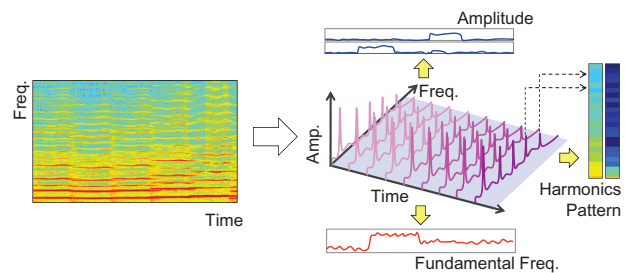


図 1: 調波パラメトリック NMF の概念図

の NMF は、演奏の第 i 音の振幅スペクトログラム $Y_{t,f,i}$ を K 個のスペクトルパターン $W_{f,k}$ とその非負係数 $U_{i,k,t}$ の線形結合で表す (t は時間フレーム, f は周波数ビン)。

$$Y_{t,f,i} = \sum_{k=1}^K W_{j,k} U_{i,k,t} \quad (W_{j,k}, U_{i,k,t} > 0) \quad (1)$$

ここで、係数 $U_{i,k,t}$ を楽譜情報ベクトルの値で固定すれば回帰式となり、実演奏音響信号の振幅スペクトログラム $X_{t,f}$ との間の何らかの距離尺度 $\sum_{t,f} D(X_{t,f} || \sum_i Y_{t,f,i})$ の最適化によりスペクトルパターン $W_{f,k}$ を推定できる。

しかし、1) 各時刻各周波数の振幅スペクトルの値は楽譜から容易に説明できるような特徴ではない、2) NMF の特性上、基本周波数の異なる楽器音やビブラートのかかった演奏を表現しにくい、といった問題を持つため、この回帰モデルの設計には改善点がある。

3. 調波パラメトリック NMF

前章を踏まえ、新しいスペクトル表現法「パラメトリック NMF」及びその特殊形「調波パラメトリック NMF (HPNMF)」の定義について述べる。

パラメトリック NMF では、周波数軸上に定義された J 個の核関数 $\phi_{j,f}^{t,i}$ の線形和で振幅スペクトルをモデル化し、さらにその結合重みを NMF と同様に分解する。

$$Y_{t,f,i} = \sum_j \phi_{j,f}^{t,i} \left(\sum_{k=1}^K W_{j,k} U_{i,k,t} \right) \quad (2)$$

j は核関数のインデックスである。これはもはや振幅スペクトログラムの NMF ではないが、モデルが表現する対象は依然として振幅スペクトログラム自体である。

パラメトリック NMF の最大の利点は、推定対象の特性に合わせて核関数を定義することで、音高・音色などの音響特徴を明示的に扱えることである。今回の場合図 1 のように、周波数軸に整数倍に並んだ調波構造用の鋭いガウス関数群と、拡がりの大きい非調波構造用ガウス関数群の 2 要素 [1, 3] で調波楽器音をモデル化する。

$$\phi_{j,f}^{t,i} = \begin{cases} \exp\left(-\frac{(\omega_f - j\mu_{t,i})^2}{2\sigma^2}\right) & (j = 1, \dots, H) \\ \exp\left(-\frac{(\omega_f - v_j)^2}{2\gamma^2}\right) & (j = H + 1, \dots, J) \end{cases} \quad (3)$$

ω_f は周波数ビン f に対応する周波数 (Hz) であり、 σ^2, γ^2 及び $\{v_j\}_{H+1 \leq j \leq J}$ は定数である。 σ^2 は調波構造ピークの周波数方向拡がりに合わせて設定し、 γ^2 は大きな値を設定する。 v_j は周波数軸上で均等に並べる。調波構造用の核関数 $\{\phi_{j,f}^{t,i}\}_{1 \leq j \leq H}$ は動的であり、F0 パラメータ $\mu_{t,i}$ を各時刻、各単音毎に与えることで、ビブラート等 F0

表 1: 本稿で用いた楽譜情報ベクトルの構成法.

1. 各単音毎に決まる以下の情報を算出し 0 から 1 に規格化: -音高(ノートナンバー), -前音との音高差, -後音との音高差, -2つ前の音との音高差, -音高に対応する周波数, -音長, -前音の音長, -後音の音長, -2つ前の音の音長
2. 各時間フレーム毎に, 各単音中の開始位置からの相対時間フレームを算出し 0 から 1 に規格化
3. 1 と 2 の情報の各組み合わせを, 平均の異なる複数の 2 次元ガウス基底関数で変換し, 結果を $U_{i,k,t}$ に割り当てる.

が変化する音を効果的にモデル化できる*. この核関数構成によるパラメトリック NMF を HPNMF と呼ぶ.

HPNMF のパラメータは $\{\mu_{t,i}, W_{j,k}, U_{i,k,t}\}$ となる. ただし本稿の目的は楽譜からの振幅スペクトル生成なので, $U_{i,k,t}$ は楽譜情報ベクトルで固定し回帰モデルとする.

4. HPNMF による回帰

4.1 楽譜情報ベクトルの設計

ピアノ演奏の音量・タイミング情報生成の研究 [4] を参考に, Standard MIDI File から容易に取得可能な範囲の情報から表 1 の方法で楽譜情報ベクトルを構成する. 多項式回帰ではなくガウス基底回帰とした理由は, 説明変数の影響範囲を局所化するためである. NMF を回帰に用いる場合, 説明変数 $U_{i,k,t}$, 回帰係数 $W_{j,k}$ がともに非負の範囲に限定されるので, ある要因によって広範囲の振幅スペクトルに大きな値が与えられると, 他の要因が入り込む余地がなくなり十分な性能を示さなくなる. この問題を避けるには, 例えば音高と発音時刻からの相対時間がそれぞれ特定の値近辺にあったときのみ大きな値となるような説明変数の変換写像が必要であり, この写像にガウス関数を用いる.

また F0 $\mu_{t,i}$ については, 推定時は各単音・各時刻の値を直接推定し, 推定を終えた後で回帰を行う. その理由は, 1) F0 の推定誤差は振幅スペクトルの推定精度に強く影響するから, 2) ピブラートをうまく説明する説明変数, 基底関数の設計が困難であるから, の 2 点である.

4.2 パラメータ推定

モデル $Y_{t,f} = \sum_i Y_{t,f,i}$ と実演奏音響信号 $X_{t,f}$ の間の距離尺度 D として音源分離問題によく用いられる I ダイバージェンス [5] を採用し, これを最小化する.

$$\text{minimize} \sum_{t,f} \left(X_{t,f} \log \frac{X_{t,f}}{Y_{t,f}} - (X_{t,f} - Y_{t,f}) \right) \quad (4)$$

上式を最小化する $\mu_{t,i}$ と $W_{j,k}$ は解析的に得られないので, 補助関数法 [2, 3] と呼ばれる反復更新アルゴリズムを用いる. 上式第 1 項は $Y_{t,f}$ から見て逆数の対数, すなわち凸関数であり, 以下の Jensen の不等式が成り立つ.

$$\log \frac{1}{Y_{t,f}} \leq \sum_{j,i,k} \lambda_{i,j,k}^{t,f} \left(\log \frac{\lambda_{i,j,k}^{t,f}}{\phi_{j,f}^{t,i} W_{j,k} U_{i,k,t}} \right) \quad (5)$$

ただし $\lambda_{i,j,k}^{t,f}$ は $\forall t, f, i, j, k: \lambda_{i,j,k}^{t,f} > 0, \forall t, f: \sum_{i,j,k} \lambda_{i,j,k}^{t,f} = 1$ を満たす補助変数である. 式 (4) の第 1 項の $\log(1/Y_{t,f})$ をこの不等式の右辺で置き換えると

$$\sum_{t,f,i,j,k} \left(\lambda_{i,j,k}^{t,f} X_{t,f} \log \frac{\lambda_{i,j,k}^{t,f}}{\phi_{j,f}^{t,i} W_{j,k} U_{i,k,t}} + \phi_{j,f}^{t,i} W_{j,k} U_{i,k,t} \right) \quad (6)$$

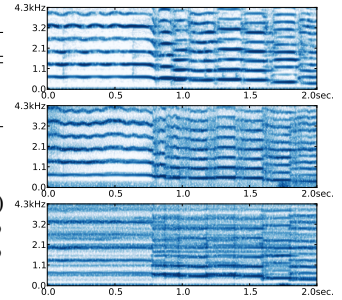
となり (ただし $\mu_{t,i}, W_{j,k}$ に非依存の項を省略), これを補助関数と呼ぶ. 証明は省略するが, 式 (6) を補助変数及

*非調波構造用の核関数 $\{\phi_{j,f}^{t,i}\}_{H+1 \leq j \leq J}$ は表記上添字 t, i を備えているが全時間フレーム, 全単音で不変な関数とする.

表 2: 演奏合成結果の SNR

	HPNMF	NMF
Violin	2.79 dB	2.40 dB
Cello	2.77 dB	2.51 dB
Flute	2.02 dB	1.28 dB

図 2: スペクトログラム.(上) 実演奏, (中) HPNMF による合成結果, (下) NMF による合成結果



びパラメータに対し交互に最小化することで元の式 (4) の局所最適解が得られる. 式 (6) の局所最小値を与える各更新式は以下のとおり解析的に導出できる.

$$\lambda_{i,j,k}^{t,f} \leftarrow \frac{W_{j,k} U_{i,k,t} \phi_{j,f}^{t,i}}{Y_{t,f}} \quad (7)$$

$$W_{j,k} \leftarrow \frac{\sum_{t,f,i} \lambda_{i,j,k}^{t,f} X_{t,f}}{\sum_{t,f,i} U_{i,k,t} \phi_{j,f}^{t,i}}, \quad \mu_{t,i} \leftarrow \frac{\sum_{j=1}^H j \sum_{f,k} \omega_f \lambda_{i,j,k}^{t,f} X_{t,f}}{\sum_{j=1}^H j^2 \sum_{f,k} \lambda_{i,j,k}^{t,f} X_{t,f}} \quad (8)$$

なお式 (6) 第 2 項は $\mu_{t,i}$ に非依存という近似を用いた [3].

5. 評価実験

HPNMF を用いた楽器演奏の分析合成の有効性を示すために行った動作実験について述べる. 市販 CD 収録のプロによる無伴奏単旋律演奏: Violin, Flute, Cello 各 3 曲の計 9 曲に対し, 各曲の後ろ 4/5 の演奏音響信号と楽譜を用い演奏を分析し, その推定結果から前 1/5 の楽譜に対する演奏の振幅スペクトログラムを合成する. 提案法との比較のために通常の NMF による分析合成も行い, HPNMF との挙動の違いを調べる. 合成結果と元の実演奏との間の信号対雑音比 (SNR) の大きさを評価する.

表 2 に演奏合成結果の SNR の楽器毎の平均を, 図 2 には合成結果例のスペクトログラムを示す. SNR による評価では HPNMF と NMF は僅かな差であったが, 聴感上の差は大きく, HPNMF による合成結果には調波構造が明確に感じ取れるのに対し, NMF による合成結果には調波構造が不明瞭な箇所があった. その違いはスペクトログラムからも見て取れる. 現在の合成品質は十分とは言えないが, 改良案として, 音楽的知見からより妥当な回帰モデルを設計する, モデルが示す振幅スペクトログラムを補正する [1], といった展望がある.

6. おわりに

本稿では HPNMF という新しい楽器音モデリング手法を用いた楽器演奏分析合成法を報告した. 今後は, 回帰式の設計について精査し演奏合成性能を向上させる他, 伴奏や残響が存在するようなより困難な条件下での本手法の挙動を精査する. なお本研究は科研費 (S), グローバル COE, CrestMuse の支援を受けた.

参考文献

- [1] 安良岡 他: 多重奏音響信号中の演奏をユーザー指定の旋律に差し替えるフレーズ置換システム. 情処研報, 2010-MUS-86, 2010.
- [2] Lee et al.: Algorithms for Non-negative Matrix Factorization. In Proc. NIPS, pp. 556-562, 2000.
- [3] Kameoka et al.: Speech Spectrum Modeling for Joint Estimation of Spectral Envelope and Fundamental Frequency. IEEE Trans. Audio, Speech and Lang. Process., Vol. 18, No. 6, pp. 1507-1516, 2010.
- [4] Teramura et al.: Gaussian Process Regression for Rendering Music Performance. In Proc. ICMPC, pp. 167-172, 2008.
- [5] FitzGerald et al.: On the Use of the Beta Divergence for Musical Source Separation. In Proc. ISSC, pp. 1-6, 2009.