

# 音楽信号の多重音解析と音色特性のデジタルフィルターの推定

日下部 峻<sup>†</sup> 久保田 光一<sup>‡</sup>

中央大学大学院 理工学研究科 情報工学専攻<sup>†‡</sup>

**概要:** シンセサイザーのパラメーターを設定し、音色の再現を行うことは、初心者にとって困難な作業である。そこで、シンセサイザーのパラメーター設定から合成音を計算する問題を順問題とした時の逆問題として、既存音源を元にシンセサイザー音のパラメーターを推定、提示し、音色の再現を支援するアルゴリズムを提案する。Nonnegative Matrix Factor 2D Deconvolution (NMF2D)[2] を用いることで、頻出するスケールパターン行列を抽出し、そこからヴァーチャルアナログシンセサイザーのフィルターセクションのカットオフ周波数包絡を推定する。

**キーワード:** 多重音解析, 非負値行列因子分解, シンセサイザー

## 1 はじめに

ポピュラーミュージックの楽曲制作において、シンセサイザーは欠かせないツールである。しかし、初心者にとって、シンセサイザーのパラメーターを調整し、思い通りの音色を出力することは難しい。

そこで、音楽信号から周波数特性を解析し、減算合成方式シンセサイザーのフィルターセクションのパラメーターを提示する方法を提案する。また、多重音解析によって既存楽曲から音色の周波数成分を抽出し、提示されるシンセサイザーのパラメーターの近似値を手がかりにすることによって、シンセサイザーの操作の難易度を下げたい。

非負値行列因子分解 (Nonnegative Matrix Factorization; NMF)[1] の拡張手法である NMF2D は音色の調波構造が対数周波数軸上でシフト不変であると仮定され、スペクトログラムパーツの対数周波数・時間平面での 2 次元畳み込みで楽音スペクトログラムを表現する。音色ごとに基底を集約することができ、NMF では困難だった周波数の時間変化を扱える。

まず、単音の信号からフィルターエンベロープを推定する方法を提示する。そして、NMF2D を用いることによって多重音信号からもフィルターエンベロープを推定する。

## 2 Synth1 のモデル化

本研究ではソフトウェアシンセサイザーの Synth1[3] をモデリング対象とする。Synth1 のフィルターとアンプリファイアのエンベロープパラメーターを推定するために、パラメーターの値を調べた。

Synth1 のエンベロープはアタックタイム、ディケイタイム、サステインレベル、リリースタイムを操作する 4 つのパラメーター A, D, S, R で構成され、0 から 127 の離散値をとる。

### 2.1 継続時間

A, D, R のパラメーター値によって、継続時間がどの程度長くなるかを調べた結果が表 1 である。アタックタイムは D, S,

表 1 パラメーターの変化と継続時間 [s] の変化

A, D, R	0	32	64	96	127
アタックタイム	0.211	0.220	0.336	2.153	28.2
ディケイタイム	0.211	0.272	0.822	5.73	46.4
リリースタイム					

R の値を 0 にし、A のみを変化させた。ディケイタイムは A, S, R の値を 0 にし、D のみを変化させた。リリースタイムは A, D の値を 0, S の値を 127 にし、R のみを変化させた。ディケイタイムとリリースタイムの結果は同様になった。

この結果より、最短発音時間は 0.211[s] であることがわかる。平均レベルがとても小さいため、本研究では考慮しない。

観測結果から、最短発音時間を引いたものが表 2 である。表 2 のアタックタイムより、パラメーター A が 32 増えるごとに

表 2 パラメーターの変化と継続時間 [s] の変化

A, D, R	32	64	96	127
アタックタイム	0.009	0.125	1.94	28.0
ディケイタイム	0.061	0.611	5.51	46.2
リリースタイム				

継続時間がおおよそ 16 倍になることがわかる。アタックタイムのおおよその継続時間は式 (1) で近似される。

$$T_a = 0.000479 \cdot (2)^{\frac{a}{8}} \quad (1)$$

式 (1) において、 $T_a$  はアタックタイムの継続時間 [s],  $a$  はパラメーター A に対応する値である。

同様に、表 2 のディケイタイムより、D のパラメーターが 32 増えるごとに継続時間がおおよそ 9 倍になることがわかる。ディケイタイムのおおよその継続時間は式 (2) で近似される。

$$T_d = 0.00755 \cdot (3)^{\frac{d}{16}} \quad (2)$$

式 (2) において、 $T_d$  はディケイタイムの継続時間 [s],  $d$  はパラメーター D に対応する値である。リリースタイムについても、式 (2) を適用する。

本研究ではアタックタイムの継続時間のモデルを式 (1)、ディケイ・リリースタイムの継続時間のモデルを式 (2) とする。

### 2.2 レベル変化

アタックタイムのレベルの変化は線形で近似できる。レベルの最大値を  $M$  とすると、時間  $t$  のレベル  $l_a$  は式 (3) とする。

$$l_a(t) = \frac{M}{T_a} t \quad (3)$$

サステインレベルはパラメーター S によって与えられる。対応するパラメーター値  $s$  による出力レベル  $l_s$  は式 (4) とする。

$$l_s(t) = \frac{M}{127} s \quad (4)$$

ディケイタイムとリリースタイムのレベルの変化は、同様の傾向が見られた。アタックタイムと同様、時間  $t$  のレベル  $l_d$  は式 (5) とし、レベル  $l_r$  は式 (6) とする。T は全体の長さである。

$$l_d(t) = (M - l_s(T_d)) \exp\left(\frac{-8}{T_d}(t - T_a)\right) + l_s(T_d) \quad (5)$$

$$l_r(t) = l_d(T - T_r) \exp\left(\frac{-8}{T_r}t - (T - T_r)\right) \quad (6)$$

以上をまとめると、時間  $t$  におけるレベル  $l$  は式 (7) になる。

$$l(t) = \begin{cases} \frac{M}{T_a} t & [0 \leq t \leq \min(T_a, T - T_r)] \\ (M - \frac{M}{127}s) \exp\left(\frac{-8}{T_d}(t - T_a)\right) + \frac{M}{127}s & [T_a < t \leq \min(T_a + T_d, T - T_r)] \\ \frac{M}{127}s & [T_a + T_d < t \leq T - T_r] \\ ((M - \frac{M}{127}s) \exp\left(\frac{-8}{T_d}(T - T_r - T_a)\right) + \frac{M}{127}s) \exp\left(\frac{-8}{T_r}(t - (T - T_r))\right) & [T - T_r < t \leq T] \end{cases} \quad (7)$$

## 3 NMF2D

NMF2D はモノラル音源分離の手法として注目されている NMF の拡張手法である。[2]

Multiple sound analysis of the music signal and Digital filter estimation of acoustic characteristics

<sup>†</sup> Takashi KUSAKABE, Information and System Engineering Course, Graduate School of Science and Engineering, CHUO University

<sup>‡</sup> Koichi KUBOTA, Information and System Engineering Course, Graduate School of Science and Engineering, CHUO University

3.1 定式化

NMF2D は観測行列  $Y$  を基底行列  $W^\tau$  と係数行列  $H^\phi$  の畳み込みで表現する.

$$Y \approx X = \sum_{\tau=0}^{T-1} \sum_{\phi=0}^{\Phi-1} W^{\tau} H^{\phi} \quad (8)$$

上式において  $X$  を再構成行列と呼ぶ. 行列  $Y, X, W, H$  の要素をそれぞれ  $y_{i,j}, x_{i,j}, w_{i,k,\tau}, h_{k,j,\phi}$  とすると, 式 8 は以下のように書き下せる.

$$y_{i,j} \approx x_{i,j} = \sum_{k=0}^{K-1} \sum_{\tau=0}^{T-1} \sum_{\phi=0}^{\Phi-1} w_{i+\phi,k,\tau} h_{k,j-\tau,\phi} \quad (9)$$

NMF2D も NMF と同様に, 観測行列と再構成行列の間に近似誤差の乖離度を非負値制約のもとで定義し, 最小化する. 本研究では, Euclid 距離の二乗誤差を用いる.

$$D_{EU}(y_{i,j}|x_{i,j}) = (y_{i,j} - x_{i,j})^2 \quad (10)$$

3.2 補助関数

式 (10) を変数  $x$  に関して微分し,  $D'_{EU}(y|x) = 0$  となる  $x$  を求めたい. 非負制約を保ったまま解析的に解くことができない. そこで, 目的関数の上限となる補助関数を設計し, 補助関数を降下させることで目的関数を間接的に降下させる.

3.3 乗法更新式

Jensen の不等式を用いて補助関数を設計することで以下の乗法更新則を得る ( $\circ$  は成分ごとの積である).

$$W^\tau = W^\tau \circ \frac{\sum_{\phi} \downarrow_{\phi \rightarrow \tau} Y H^\phi}{\sum_{\phi} \downarrow_{\phi \rightarrow \tau} X H^\phi} \quad (11)$$

$$H^\phi = H^\phi \circ \frac{\sum_{\tau} \uparrow_{\tau \leftarrow \phi} W^\tau Y}{\sum_{\tau} \uparrow_{\tau \leftarrow \phi} W^\tau X} \quad (12)$$

4 提案法

音源信号から, Synth1 のパラメーターを探索する. 音源解析を行い, 式 (7) との二乗誤差が小さくなるようにパラメーターを推測する.

- Step1. NMF2D を用いて基底行列  $W^\tau$  を求める.
- Step2.  $W^\tau$  の逆ウェーブレット変換で得られる波形の概形との二乗誤差が小さくなるようにアンプリファイアーエンベロープを探索する.
- Step3.  $W^\tau$  から基音を調べる.
- Step4.  $W^\tau$  の基音と同じ周波数の Sawtooth 波に Step2 で推定したエンベロープを掛け,  $w'^\tau$  とする.
- Step5.  $w'^\tau$  の Wavelet 変換  $W'^\tau$  にフィルターをかける.
- Step6.  $W$  と  $W'^\tau$  の二乗誤差が小さくなるようにフィルターエンベロープを探索する.

5 評価実験

5.1 単音の解析

Synth1 によって出力した単音データを解析し, パラメーターを推測する. 推測されたパラメーターと実際のパラメーターを比較する.

今回の実験に用いた音源信号は Synth1 のプリセットである Piano を 0.75 秒間 MIDI ノートによって出力し, 録音した音声ファイルである.

Piano のアンプリファイアーのエンベロープの推定結果が表 3 である. また, 推定されたエンベロープを図示したものが図 1 である.

表 3 で D の値が最も異なっている. 今回モデリングしていないオシレーターの影響によって振幅が揺らいでいることと, リリースタイムの探索との兼ね

合いで十分な長さの MIDI ノートで出力していないことが原因として考えられる.

また, フィルターのエンベロープの推定結果が表 4 である. 図 2 は実験に用いた音源信号を Wavelet 変換した結果であり, 図 3 は推定されたアンプリファイアーとフィルターのエンベロープから合成したスカラグラムである.

表 4 で R の値が大きく異なっているのは, Wavelet 変換は低い周波数ほど時間方向に分解能が落ちてしまうことや, フィルターのリリースタイムよりアンプリファイアーのリリースタイムの方が短いため, 信号の減衰によってフィルターのリリースタイムの途中で信号が失われてしまったことが考えられる.

表 3 プリセット Piano のアンプリファイアーのエンベロープ

	A	D	S	R
Original	0	97	0	47
Estimate	0	89	0	40

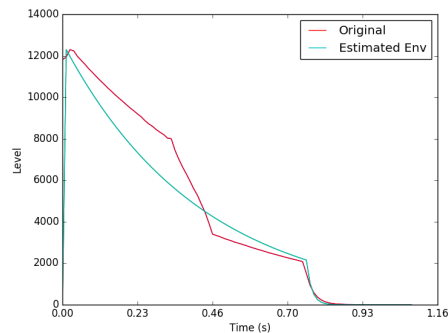


図 1 プリセット [Piano] のアンプリファイアーのエンベロープ

表 4 プリセット Piano のフィルターのエンベロープ

	A	D	S	R
Original	0	73	73	62
Estimate	0	68	99	74

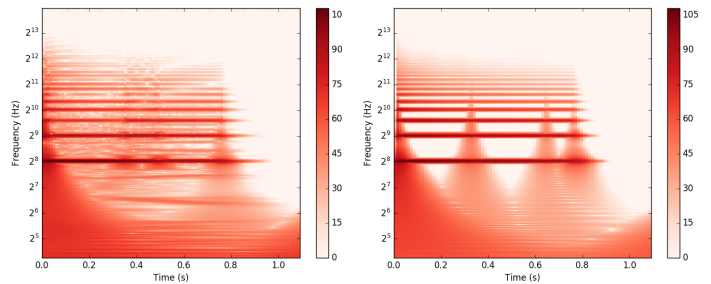


図 2 プリセット Piano のスカラグラム

図 3 推定したエンベロープから合成したスカラグラム

6 今後の課題

Synth1 によって出力した多重音データを解析し, パラメーターを推測する. 推測されたパラメーターと実際のパラメーターを比較する.

参考文献

- [1] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," Advances in Neural and Information Processing Systems 13, pp.556562, 2001.
- [2] M. Mørup and M. N. Schmidt, "Sparse non-negative matrix factor 2-D deconvolution," Technical Report, Technical University of Denmark, 2006.
- [3] "Daichi Laboratory," <http://www.geocities.jp/daichi1969/>, 最終アクセス日 2015-12-16.