

# 確率スペクトル包絡に基づく楽器音の音源同定

A Musical Instrument Identification Method Based on Probabilistic Spectrum Envelopes

辰巳 百絵子† 上野 敦志† 田窪 朋仁†

† 大阪市立大学大学院工学研究科

## 1 はじめに

近年、様々な音楽アプリケーションに応用できる基礎技術として、楽音解析の研究が重要視されている [1]。しかし混合楽音解析は十分な精度が得られておらず、楽器推定もその課題の 1 つとなっている [2]。本研究では、教師データとして音階を鳴らした音源のみを用い、楽器カテゴリ毎の特徴を確率スペクトル包絡 (Probabilistic Spectrum Envelope; PSE) で表現し、テストデータの PSE を各楽器の PSE と比較して楽器音の音源同定を行う手法を提案する。

## 2 音源同定

### 2.1 音源同定の流れ

まず音響信号に短時間フーリエ変換 (Short-Time Fourier Transform; STFT) と非負値行列因子分解 (Non-negative Matrix Factorization; NMF) を行って、各音の周波数特徴を抽出し、ガウシアンカーネルを用いて PSE を表現する。次に PSE を比較して音源同定を行う。音源同定のフローチャートは図 1 のようになる。

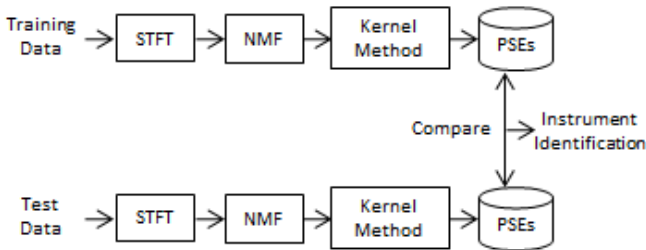


図 1: 音源同定のフローチャート

### 2.2 非負値行列因子分解 (NMF)

音響信号を短時間に分割して STFT を行い、周波数に対する振幅の値を求める。それらの値を列ベクトルとし、全時間の列ベクトルの集合を、振幅スペクトログラム  $\mathbf{V}$  とする。次に教師なし NMF を実行し  $\mathbf{V} \approx \mathbf{WH}$  に分解する [3]。基底行列  $\mathbf{W}$  は各音の周波数特徴を表した行列で、アクティビティ行列  $\mathbf{H}$  は各音の各時間での強さを表した楽譜のような行列になっている。このうち基底行列  $\mathbf{W}$  の各列からの極大値となる点を取り出し、そのうちの上位 20 点ずつをピーク点とする。ピーク点集合  $\{(f_n, y_n)\}$  は各点での周波数と振幅の値のセットとした集合になっている。

### 2.3 PSE のカーネル表現

PSE は  $p(f, y)$  と表され、ある周波数  $f$  で切った断面図は、確率密度関数になっている。まず、周波数領域に関して、先ほど求めたピーク点集合の周波数の値を平均、 $\sigma_1^2$  を分散としたガウシアンカーネルを求

め、周波数の重みとする。次に振幅領域に関しても同様に、各ピーク点の振幅の値を平均、 $\sigma_2^2$  を分散としたガウシアンカーネルを求める。最後にそれら 2 つを組み合わせる PSE を表現する。  $p(f, y)$  は以下の式のようになる。  $A$  は正規化係数である。

$$p(f, y) = A \sum_{n=1}^N \left[ \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left\{ -\frac{1}{2} \left( \frac{f - f_n}{\sigma_1} \right)^2 \right\} \right] \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp \left\{ -\frac{1}{2} \left( \frac{y - y_n}{\sigma_2} \right)^2 \right\}$$

## 3 実験と結果

音源データは MuseScore 2.0.2 を用いて作成した。教師データは、ピアノ C2~B5(48 音)、バイオリン G3~B6(41 音) をそれぞれ個別に鳴らした音階とした。テストデータはクラシック音楽 5 曲から一部を抜粋し、ピアノ、バイオリン、ピアノとバイオリンの合奏の 3 パターンで演奏したものとした。計 15 曲のテストデータについてどのパターンの演奏かを判定する実験を 1 回とした。NMF の初期値をランダムで与えて 10 回実験を行った結果は表 1 となった。バイオリン演奏の誤判定は、全て合奏と判定されており、合奏判定を除いた場合の正解率は 100% となっていた。

	Pf	Vn	P+Vn	全平均
正解率 (%)	100	66.0	100	88.0

表 1: 各楽器の正解率比較

## 4 結論・今後の課題

PSE を用いて楽曲中の楽器音の音源同定を行い、正解率は 88.0% となった。今後は、同一の楽器カテゴリに含まれる楽器の PSE を比較し、その音源同定の精度を確かめていく。

### 参考文献

- [1] 亀岡弘和, 中村友彦, 高宗典玄, "音楽音響信号処理技術の最先端," 電子情報通信学会誌, vol. 98, no. 6, pp. 467-474, 2015.
- [2] E. Kubera, and A. Wiczorkowska. "Mining audio data for multiple instrument recognition in classical music," New Frontiers in Mining Complex Patterns. Springer International Publishing, pp.246-260, 2013.
- [3] D.Lee, H.Seung, "Algorithms for non-negative matrix factorization," Advances in neural information processing systems, vol.13, pp.1-7, 2001.