

# 楽曲中の打楽器と調波楽器の音を分離する

宮本 賢一 角尾 衣未留 嵯峨山 茂樹

(嵯峨山グループ：東京大学情報理工学系研究科システム第一研究室)

**概要** 本研究は、ポピュラー音楽などの1ch音楽信号から、調波的な楽器音成分と打楽器的な楽器音成分を分離するものである。この分離手法は、打楽器やノイズを含んだ多声音楽信号の楽音分析における前処理、打楽器パートの強調や打楽器パターン変更といった音楽加工など、多くの応用が期待される。特に本研究では、調波音と打楽器音におけるスペクトログラムの滑らかさの異方性に基づき、EMアルゴリズムを利用した観測スペクトログラムを両者に分離する時間周波数マスクの反復推定手法を用いて、実時間分離システムを実現した。

**キーワード**：信号処理、スペクトログラム、分離、EMアルゴリズム。

## 1. はじめに

近年音楽音響信号処理の研究分野において、単一チャンネル多声音楽信号からのピッチ・和音推定、リズム推定など様々な分析技術が開発されているが、ポピュラー音楽など、音程を持つ楽器音と非調波的な打楽器音が混合された音楽信号においては、これらの分析は難しいと考えられている。そこで、1ch音楽信号から調波的な楽器音成分と打楽器的な非調波音成分を分離する手法を提案する。この分離は、打楽器やノイズなどの非調波成分を含んだ多声音楽信号の楽音分析における前処理、打楽器パートの強調や打楽器パターン変更といった音楽加工など、多くの応用が期待される。

関連研究としては、各フレームにおいて周期性・非周期性の性質を用いた成分分離を行なう手法 [1]、除去対象の打楽器のスペクトルテンプレートを用いた打楽器同定・除去手法 [2]、分析対象楽曲の楽譜情報 (MIDI 情報) を用いた調波・非調波構造のモデルによる楽音分離手法 [3] などが挙げられる。

それに対し我々は、楽器や楽譜に関する情報を全く用いずに、単一チャンネル音楽信号からの分離手法として、スペクトログラム上で画像処理的な2次元フィルタを用いた高速な直接計算手法 [4] を開発した。本研究ではさらに性能を向上させた、スペクトログラムの滑らかさの異方性に基づいたEMアルゴリズムによる反復解法 [5] を適応し、実時間で分離するシステムを提案する。

## 2. スペクトログラムの分解

本研究では調波音と打楽器音の混在した 1ch 音楽信号を分析対象とし、入力信号の短時間周波数解析によって得られるスペクトログラムを  $W(x, t)$  とする ( $x$ : 周波数,  $t$ : 時刻)。本研究の問題は、この  $W(x, t)$  を打楽器的な音程を持たない非調波成分

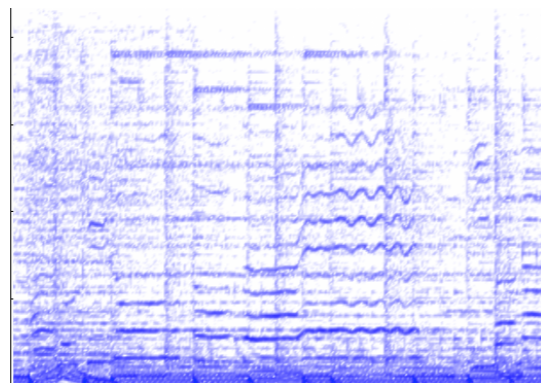


図1 ポピュラー音楽のスペクトログラム

$P(x, t)$  と音程を持つ楽器のような調波成分  $H(x, t)$  の2つのスペクトログラムに分解することと考えられる。

### 2.1 着眼点：調波成分・打楽器成分の異方性

本研究では、図1で示すようなポピュラー音楽の音響信号のスペクトログラムが、一般的に周波数方向に形成される山脈と時間方向に形成される山脈とからなることが多い点に着目する。前者は、打楽器のように時間方向には急峻に変化するが周波数方向にはブロードである成分  $P(x, t)$  に、後者は逆に周波数方向には急峻な形状だが時間方向には滑らかな成分  $H(x, t)$  に対応するとみなすことができ、また2成分は時間周波数平面上においてスパースに存在しているとみなせる。

### 2.2 目的関数の最小化によるパラメータ反復推定

観測スペクトログラムを調波成分・打楽器成分に分配する時間周波数マスク  $m_H(x_i, t_j)$ ,  $m_P(x_i, t_j)$  ( $\forall i, j; m_H(x_i, t_j) + m_P(x_i, t_j) = 1$ ) を導入し、分配された

エネルギー分布  $m_P(x_i, t_j)W(x_i, t_j)$ ,  $m_H(x_i, t_j)W(x_i, t_j)$  と  $P(x_i, t_j)$ ,  $H(x_i, t_j)$  との近さを表す分布間距離として  $I$ -Divergence を採用する。

本研究では、スペクトログラムの滑らかさの異方性を、最小化すべきコストとして、隣り合う時間周波数 bin とのエネルギーの平方根の二乗誤差とし、分布間距離と滑らかさコストとの和による目的関数を最小化する問題として定式化できる。

この目的関数から、時間周波数マスクを固定して目的関数を最小化する  $H(x_i, t_j)$  と  $P(x_i, t_j)$  の更新と、 $H(x, t)$ ,  $P(x, t)$  を固定して目的関数を最小化するような  $m_P(x_i, t_j)$  と  $m_H(x_i, t_j)$  の更新を交互に行なうことにより、目的関数の最小化における局所最適解が得られる (更新式は省略)。

### 2.3 実時間分離システムの実現

前節で提案した解法は、入力信号全体の時間周波数領域における反復解法であるため、一般的には実時間分離は難しい。しかし本稿では、滑らかさを隣接した時間周波数 bin のみを用いた微分的なコストとして表現したため、局所的な分析領域でもある程度妥当な解が得られると考えられる。そこで、局所的な分析時間区間を用い、分析区間の移動とパラメータの反復更新 (1~数回) を交互に行なうことで、実時間での調波音・打楽器音分離システムを実現した。

## 3. 提案アルゴリズムの評価

### 3.1 実際の楽曲への適応

ポピュラー音楽の実演奏信号を入力信号として、RWC 研究用音楽データベースから RWC-MDB-P-2001 No.7 より抜粋して使用した (16kHz サンプリング)。入力信号のスペクトログラムを図1に、提案アルゴリズムの分離結果を図2に示す。

結果から、 $P(x, t)$ ,  $H(x, t)$  が着目した性質を満たすように分離されたことが分かる。結果の音声を聴くと、[4]の手法に比べ良く分離でき、特に調波音は非常にスムーズに聴こえた。しかし、先行研究と同様、ハイハットやバスドラムの duration 部分が  $H(x, t)$  に分離されること、歌声のビブラートや子音が  $P(x, t)$  に分離されやすいことを確認した。

### 3.2 考察

スペクトログラムの滑らかさの異方性に基づく解法が、[4]による解法と同様の性質をもった分離を、実時間に比べて十分高速にかつより高い性能で実現した。楽器の知識を用いずに簡便な特徴に基づいた解法のため、比較的音長の長いバスドラムやハイハットの打楽器音、ピアノの打鍵音、ピッチの変化しやすい歌声などは着目した特徴を満たしにくく、楽器分類の通念とは必ずしも対応しない可能性があるが、実時間演算で分離できるメリットは非常に大きいと考えられる。

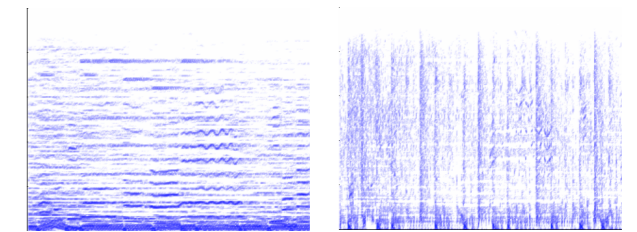


図2 調波成分と非調波成分への分離結果のスペクトログラム 左:  $P(x, t)$ , 右:  $H(x, t)$

## 4. おわりに

本研究では 1ch 音楽信号から調波的な成分と打楽器的な成分を分離する問題に対し、スペクトログラムの滑らかさの異方性に基づいた反復解法を提案し、楽曲への適用やパートに分かれた信号を用いた定量実験を行ない、その性能の評価を行なった。今後は、滑らかさの程度を示すパラメータの自動決定などが検討課題である。

## 参考文献

- [1] 亀岡弘和, 後藤真孝, 嵯峨山茂樹, “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.
- [2] 吉井和佳, 後藤真孝, 奥乃博, “実世界の音楽音響信号に対するドラムスの音源同定を利用したドラムイコライズシステム INTER:D の開発,” 第3回情報科学技術フォーラム FIT2004, 2004.
- [3] K. Itoyama, M. Goto et al., “Integration and Adaptation of Harmonic and Inharmonic Models for Separating Polyphonic Musical Signals,” Proc. ICASSP, 2007.
- [4] 宮本賢一, 立藪真理, ルルージョナトン, 亀岡弘和, 小野順貴, 嵯峨山茂樹, “スペクトログラム2次元フィルタによる調波音・打楽器音の分離,” 日本音響学会秋季研究発表会講演集, pp.825-826, Sep, 2007.
- [5] 宮本賢一, 亀岡弘和, 小野順貴, 嵯峨山茂樹, “スペクトログラムの滑らかさの異方性に基づいた調波音・打楽器音の分離,” 日本音響学会春季研究会発表会講演集, pp. 903-904, May, 2008.