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*

◎宮本 賢一, 亀岡 弘和†, 小野 順貴, 嵯峨山 茂樹 (東大情報理工)

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。そこで本研究では、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について議論する。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。本稿では性能の向上を目標として、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を提案し、計算時間や性能の評価を行う。また、このアルゴリズムを応用して実時間で分離するシステムを提案する。

2 問題設定: 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 $P(x,t)$ 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。このとき満たすべき要件は、任意の時間周波数 (x,t) において

$$P(x,t) \geq 0 \quad (1)$$

$$H(x,t) \geq 0 \quad (2)$$

$$P(x,t)+H(x,t) = W(x,t) \quad (3)$$

が成り立つことである。

3 本研究のアプローチ

3.1 着眼点: 調波成分・打楽器成分の異方性

前述の問題設定に対して本研究では、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。

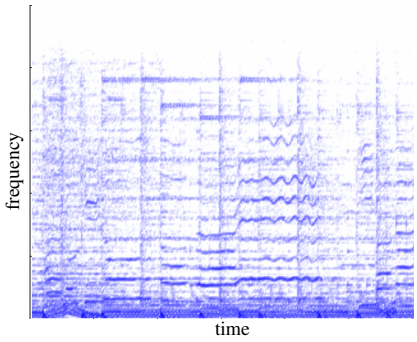

Fig. 1 ポピュラー音楽のスペクトログラムの例

3.2 滑らかさコストの導入

前節で述べたようなスペクトログラムにおける調波的な成分と打楽器的な成分の異方性を利用して、$W(x,t)$ から $H(x,t)$ と $P(x,t)$ を推定する問題を議論する。実装上 (x,t) は離散的な座標として取得できるため、以下の議論では離散的な時間周波数領域 (x_i,t_j) と定義して議論を行なう (I: 周波数 bin 数、J: 分析フレーム数)。

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

$$\Omega_H = \frac{1}{2\sigma_H^2}\sum_{i=1}^{I}\sum_{j=1}^{J-1}\left(\sqrt{H(x_i,t_{j+1})}-\sqrt{H(x_i,t_j)}\right)^2 \quad (4)$$

$$\Omega_P = \frac{1}{2\sigma_P^2}\sum_{i=1}^{I-1}\sum_{j=1}^{J}\left(\sqrt{P(x_{i+1},t_j)}-\sqrt{P(x_i,t_j)}\right)^2 \quad (5)$$

のように表現する。平方根を取ることにより、エネルギーを対数的に捉える人間の聴覚特性により近い滑らかさコストの定式化を実現した。

3.3 目的関数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 $m_H(x_i,t_j)$、$m_P(x_i,t_j)$ $(\forall i,j, m_P(x_i,t_j)+m_H(x_i,t_j)=1)$ を導入し、分配されたエネルギー分布 $m_P(x_i,t_j)W(x_i,t_j)$、$m_P(x_i,t_j)W(x_i,t_j)$ 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 を採用すると、式 (4)(5) の滑らかさコストとの和による目的関数

---

* "Separation of Harmonic and Non-Harmonic Sounds Based on Anisotropy in Spectrogram," by Ken-ichi Miyamoto, Hirokazu Kamoeka, Nobutaka Ono, and Shigeki Sagayama, Graduate School of Information Science and Technology, The University of Tokyo.

† 現在、NTT コミュニケーション基礎科学研究所に勤務

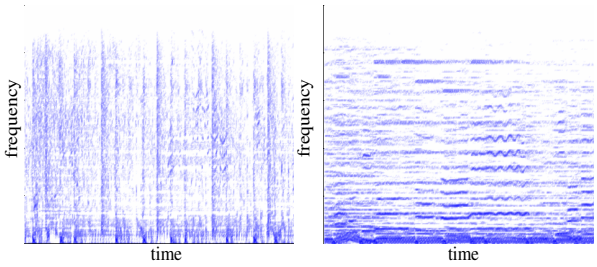

Fig. 2　調波成分と非調波成分への分離結果のスペクトログラム 左: $P(x,t)$、右: $H(x,t)$

$$\sum_{i,j} m_P(x_i,t_j)W(x_i,t_j)\log\left(\frac{m_P(x_i,t_j)W(x_i,t_j)}{P(x_i,t_j)}\right)$$
$$+\sum_{i,j} m_H(x_i,t_j)W(x_i,t_j)\log\left(\frac{m_H(x_i,t_j)W(x_i,t_j)}{H(x_i,t_j)}\right)$$
$$-\sum_{i,j}(W(x_i,t_j)-P(x_i,t_j)-H(x_i,t_j))+\Omega_H+\Omega_P \quad (6)$$

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式 (6) を最小化する $H(x_i,t_j)$ と $P(x_i,t_j)$ の更新と、$H(x,t)$, $P(x,t)$ を固定して式 (6) 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 (6) の最小化における局所最適解が得られる (更新式は省略)。

3.4　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。

4　提案アルゴリズムの評価実験

4.1　実際の楽曲への適用

本節ではポピュラー音楽の実演奏信号を用いた定性的実験を述べる。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。

4.2　パート別の分離に関する定量評価実験

次にパート別信号を用いた定量的な評価実験を行なった。RWC 研究用音楽データベースより RWC-MDB-P-2001 No.18 の前奏部 8.1 秒を入力とし、MIDI 形式データをパート別に分離し、各パートを WAV 形式に変換してその信号の和を入力とした (16kHz サンプリング)。そして [4] の手法や提案手法によって得た分離信号と各パート信号との相関を計算することで、$P(x,t)$ と $H(x,t)$ に含まれるエネルギー比率を算出し、計算時間とともに比較した (表 1、CPU3.6GHz のマシンで計算)。表より、本稿の提案手法は、[4] に比べて計算コストは増大するが、分離性能を大きく改善できることが分かる。しかし、両手法ともバスドラムは調波音側に分離された。

Table 1　パート別エネルギー分離比率

	[4] による手法		本稿の提案手法	
計算時間	0.31[s]		2.52[s]	
パート	P 比率	H 比率	P 比率	H 比率
ピアノ	0.239	0.761	0.071	0.929
ベース	0.352	0.648	0.063	0.957
シンセサイザー	0.259	0.741	0.029	0.971
E. ギター	0.233	0.767	0.074	0.926
メロディ	0.142	0.858	0.122	0.878
ブラス	0.363	0.637	0.373	0.627
スネアドラム	0.892	0.108	0.939	0.061
ハイハット	0.923	0.077	0.972	0.028
バスドラム	0.093	0.907	0.019	0.981

4.3　考察

結果より、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と言える。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。

5　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。

謝辞　本研究の一部は科学技術振興機構 CREST プロジェクトの補助を受けて行なわれた。

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湖ライズシステム 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, 立薗 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.