

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*

宮本 賢一, 立蘭 真理, ルルー ジョナトン,
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は、単一チャンネル多声音楽信号からのピッチ推定、リズム推定など様々な分析技術が開発されている。しかし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いえる。そこで本研究では、1ch 音楽信号から調波的な楽器音と打楽器のような非調波的な楽器音を分離する手法について議論する。この分離手法は、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、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]、分析対象楽曲に対応した 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。

それに対し本研究では、時間周波数平面のスペクトログラムを画像とみなし、調波的な音と打楽器的な音を持つ一般的な性質の違いを利用した2次元フィルタを用いることで、楽器固有の情報なしで1ch 音楽信号から打楽器音と調波音を分離する直接解法を提案する。

2 問題設定: スペクトログラムの分解

本研究では打楽器音と調波音の混在した1ch 音楽信号を分析対象とし、その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 $T(x, t)$ と音程を持つ楽器のような調波成分 $Y(x, t)$ の2つのスペクトログラムに分解することだといえる。このとき満たすべき要件は、任意の時間周波数において

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

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

$$T(x, t) + Y(x, t) = W(x, t) \quad (3)$$

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: 打楽器音と調波音の持つ性質

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

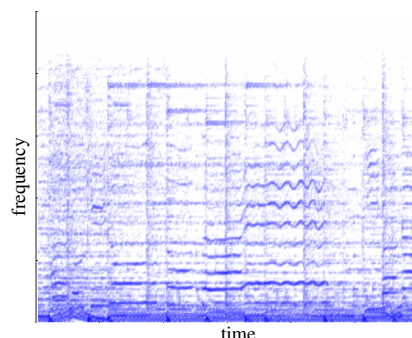


Fig. 1 ポピュラー音楽のスペクトログラムの例

化するが周波数方向にはブロードである成分 $T(x, t)$ 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 $Y(x, t)$ 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。

3.2 時間周波数マスクによる分解

前述した $T(x, t)$ と $Y(x, t)$ のスパース性から、任意の時間周波数において0か1の値をとるバイナリマスク $m_T(x, t), m_Y(x, t)$ を設計することで、

$$T(x, t) = m_T(x, t)W(x, t) \quad (4)$$

$$Y(x, t) = m_Y(x, t)W(x, t) \quad (5)$$

$$\forall x, t, \quad 1 = m_T(x, t) + m_Y(x, t) \quad (6)$$

と $W(x, t)$ を分解できると考えられる。これらの分離スペクトログラムは、式(1),(2),(3)の性質を満たす。

3.3 2次元フィルタ出力を用いたマスク設計

次に時間周波数マスク $m_T(x, t), m_Y(x, t)$ の設計について述べる。今 $W(x, t)$ を画像とみなすと、 $T(x, t)$ と $Y(x, t)$ の特徴を個別に抽出するような2次元フィルタをかけることで、そのフィルタ出力結果の大小から各時間周波数成分が $T(x, t)$ に属するか $Y(x, t)$ に属するかを決定できると考えられる。

$W(x, t)$ の2次元フーリエ変換成分を $\bar{W}(p, q)$ (p : 周波数方向のフーリエ成分, q : 時間方向のフーリエ成分) とすると、 $T(x, t)$ 特徴抽出フィルタ $\bar{F}_T(p, q)$ 、 $Y(x, t)$ 特徴抽出フィルタ $\bar{F}_Y(p, q)$ を用いることで

$$T_0(x, t) = \text{IFT}[\bar{W}(p, q) \times \bar{F}_T(p, q)] \quad (7)$$

$$Y_0(x, t) = \text{IFT}[\bar{W}(p, q) \times \bar{F}_Y(p, q)] \quad (8)$$

のようにフィルタ出力結果が得られる。この結果から時間周波数マスク $m_T(x, t), m_Y(x, t)$ は

$$m_T(x, t) = \begin{cases} 1 & (T_0(x, t) > Y_0(x, t)) \\ 0 & (\text{otherwise}) \end{cases} \quad (9)$$

$$m_Y(x, t) = 1 - m_T(x, t) \quad (10)$$

*"Separation of Harmonic and Non-Harmonic Sounds Based on 2D-filtering of the Spectrogram," by Ken-ichi Miyamoto, Mari Tatezono, Jonathan Le Roux, Hirokazu Kamoeka, Nobutaka Ono, 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

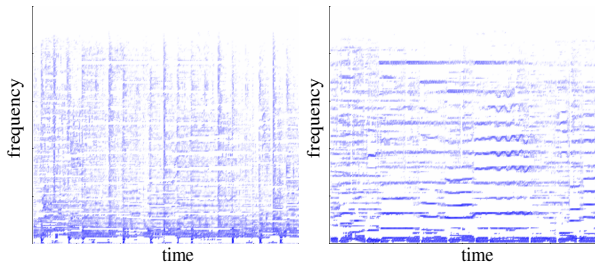


Fig. 2 分離結果のスペクトログラム 左: $T(x, t)$ 、右: $Y(x, t)$

と得られる。

3.4 特徴抽出 2次元フィルタの設計

前節で述べた2次元フィルタに関して満たすべき要件を検討する。出力結果が各時間周波数成分において $T(x, t)$ らしさ、 $Y(x, t)$ らしさの指標となるためには、フィルタ出力が非負の実数になることが望ましい。また入力スペクトログラムとフィルタ出力の時間周波数に対応している必要がある。前者の実現のためには、フィルタが任意の2次元分布の畳み込み $a(p, q) * a(p, q)$ で表現される形状であればよく、またその形状が p, q 両軸に対して線対称な実数分布になっていれば後者の性質も満たす。

3.1節で議論した $T(x, t), Y(x, t)$ の特徴を抽出する2次元フィルタ $F_T(p, q), F_Y(p, q)$ としては様々な形状が考えられる。最適なフィルタ設計については今後の検討課題であるが、本稿では前述の要件を満たす最も簡単なフィルタとして、 $F_T(p, q)$ は周波数方向のみ、 $F_Y(p, q)$ は時間方向のみのローパスフィルタ

$$F_T(p, q) = g(p) \quad (11)$$

$$F_Y(p, q) = h(q) \quad (12)$$

として設計し、 $g(p)$ や $h(q)$ の1次元ローパスフィルタ形状としては三角窓や gaussian を採用した。

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結果

本節ではポピュラー音楽の楽曲を用いた分離実験を述べる。入力信号として、RWC 研究用音楽データベースより 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 (ローパスフィルタの形状は Gaussian) による分離結果を図2に示す。

結果から、 $T(x, t)$ は周波数方向にブロードな成分、 $Y(x, t)$ は周波数方向に急峻だが時間方向に滑らかな成分に分離されたことが分かる。分離音を聴くと、スネアドラムなどの打楽器音は $T(x, t)$ に分離されたが、バスドラムやハイハットに関しては特に Duration 部分が $Y(x, t)$ に分離されることが確認された。また歌声においてピッチが連続的に変化する部分は $T(x, t), Y(x, t)$ どちらにも分離されうるが、ローパスフィルタのカットオフ周波数を調整することにより、 $Y(x, t)$ の方に多く分離することが可能であった。

4.2 MIDI を用いた定量評価実験

次に、提案アルゴリズムの定量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 の前奏部を入力とし、MIDI 形式デー

Table 1 パート別エネルギー分離比率

パート	T への分離比率	Y への分離比率
ピアノ	0.24	0.76
ベース	0.19	0.81
シンセサイザー	0.25	0.75
E. ギター	0.09	0.91
ブラス	0.38	0.62
スネアドラム	0.92	0.08
ハイハット	0.93	0.07
バスドラム	0.04	0.96

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)。そして提案手法によって得た分離結果の信号と各パート信号との相関を計算することで、 $T(x, t)$ と $Y(x, t)$ に含まれるエネルギー比率を算出した。その結果を表1に示す。表より、ギターやピアノなどのメロディーや伴奏は $Y(x, t)$ に、スネアドラムやハイハットは $T(x, t)$ に分離したが、バスドラムが $Y(x, t)$ に分離される結果を得た。

4.3 考察

本研究では、打楽器音や調波音の特徴としてスペクトログラムの周波数、時間方向の連続性を用いたが、これはスネアドラムなどの打楽器音や、音程を持つ楽器音の分離には適していると考えられる。しかし、バスドラムやハイハットのように周波数分布に偏りを持ち比較的音長の長い打楽器音や、ピアノの打鍵音やベースの打弦音、ピッチの変化しやすい歌声などはこの性質を満たさず、分離も誤りやすい結果となった。これは、今後よりよい特徴抽出2次元フィルタの形状の設計によって解決されると考えられる。

5 おわりに

本研究では1ch 音楽信号から調波的な音と打楽器的な音を分離する問題に対し、両者の性質の違いに基づいたスペクトログラム2次元フィルタを用いての直接解法を提案し、楽曲への適用例や MIDI 形式の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。今後の検討課題として、最適なフィルタ形状の考察、フィルタのパラメトリック表現における最適パラメータの自動決定などが挙げられる。

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.
- [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.
- [3] K. Itoyama, M. Goto *et al.*, “Integration and Adaptation of Harmonic and Inharmonic Models for Separating Polyphonic Musical Signals,” *Proc, ICASSP*, 2007.