

# 和声境界を考慮した単位リズムパターンの抽出に基づく 音楽音響信号の小節境界推定\*

角尾衣未留, 小野順貴, 嵯峨山茂樹 (東大・情報理工)

## 1 はじめに

近年では音楽情報検索や能動的な音楽鑑賞を目的とした音楽の自動解析に関する研究が盛んになり、楽曲を構成する重要な単位である小節へのセグメンテーションは特に需要が高い。具体的には、和声推定、ジャンル認識・音楽推奨、音響信号から MIDI(楽譜情報)へ変換など、幅広いタスクにおいて必要とされる。

従来では小節境界を推定するには主にビートの情報を用いていた [1, 2]。楽曲から打楽器のビート情報やメロディー・ベース音のオンセット情報を用いて小節境界を推定しようというアプローチである。しかしながら、ここでは 1 小節 4 拍であることを仮定しており、その他の場合には対応していなかった。それに対し我々は、楽曲は特に打楽器はある決まった複数種の小節単位パターンから構成されるという仮定のもと、その小節境界推定手法 [3] や小節単位の情報を利用したジャンル認識 [4, 5] を提案してきた。

実際には小節ごとに決められた打楽器パターンが演奏されるのみでなく、和声も小節の境界で変化しやすい特性を持っている。本稿では我々が従来提案してきた小節境界推定手法をクロマベクトルによる和声境界情報を利用することによって小節境界の推定精度を向上させる手法を提案する。

## 2 クロマベクトルを用いた和声境界推定

和声は一般に小節の境界で変化しやすい特性を持つため、和声の変わり目を検出することによって小節境界の推定の精度を上げることができると考えられる。本研究では和声境界の検出のために、多くの自動和声認識システム (例えば [6]) で特徴量として用いられているクロマベクトル [7] に着目した。

クロマベクトルはフーリエ解析によって得られる周波数領域のエネルギーをオクターブの 12 音に折り返して足し合わせたものである。図 1 の左上に典型的なスペクトログラムを示してある。これをクロマベクトルに変換した結果は図 1 の右上となる。クロマベクトルはオクターブの 12 音による表現であるため、倍音成分の影響が少なく和声の構成要素を解析しやすい。

和声が変化する時、和声を構成する要素も変化するためクロマベクトルも時間的に変化すると考えられる。しかし小節の最後の時刻までである和声内音が鳴ったまま急に次の小節で他の和声音が演奏されるとは限らず、小節の頭で鳴る音は減衰し小節の最後の時点ではエネルギーがない可能性がある。それに対し小節の頭では何らかの音が演奏されることが多い。そのため、和声境界の候補はクロマベクトルの時間

的な微分値-デルタクロマの中でも正の値を計算することによって得られると考えられる。

時刻  $t$  の正のデルタクロマ  $d(t)$  は直前の時刻のクロマベクトル  $c_i(t-1)$  と現在のクロマベクトル  $c_i(t)$  の差分を用いて

$$d(t) = \frac{1}{2} \left( \sum_{i=1}^{12} (c_i(t) - c_i(t-1)) + \left| \sum_{i=1}^{12} (c_i(t) - c_i(t-1)) \right| \right) \quad (1)$$

と表すことができる。ただし  $c_i(t)$  は時刻  $t$  の  $i$  番目のクロマベクトルのエネルギーである。図 1 の左下は図 1 の右上のクロマベクトルの正の微分値 (式 1) であり、右下は各和声境界の候補を対等に扱うために短時間 (ここではおよそ 1 秒間) で 0 から 1 の値に正規化したものである。

特に現代のポピュラー音楽ではメロディー・ベース音などの調波音以外に打楽器が同時に演奏されることが多い。実際の楽曲から上のようなクロマベクトル及び正のデルタクロマを計算するにあたり、打楽器音成分によって正しい解析が行えない可能性がある。そのため、実際の楽曲に適應する際は調波音・打楽器音分離手法 [8] を適用することによりこの問題が解決されると考えられる。この手法は打楽器音と調波音それぞれが異なる音色の特徴を一般に持つという特性を利用する：調波音は特定の周波数で比較的長時間演奏されるためスペクトログラム上では横に滑らか、逆に打楽器音は広い周波数帯域にエネルギーが広がり時間的に瞬時であるので縦に滑らかとなる。このスペクトログラムの滑らかさの異方性を利用し、MAP 推定の枠組で各成分を推定する。この手法を用いて分離強調された旋律楽器音のみを用いる事で、正のデルタクロマを計算する。

## 3 和声境界情報を利用した単位リズムパターン抽出

我々が従来提案してきた単位リズムパターン抽出手法では打楽器音のみの情報を用いていた [3]。前処理として前節で述べた調波音・打楽器音分離手法を用い打楽器音のみが用いられ、小節単位の打楽器パターンとその小節への最適なセグメンテーションが同時に推定される。初期の複数の打楽器パターンを基に動的計画法 (one-pass DP 法) でアラインメントを計算し、 $k$ -means クラスタリングに基づいてテンプレートパターンを平均計算により更新する。この二つのステップを繰り返すことにより収束後には、最適なテンプレートとセグメンテーションが得られる。

ここで小節への分割を行っている one-pass DP 法では各時間フレームとテンプレートとの距離 (もしくは

\*Musical Bar Line Estimation Based on Extraction of Unit Rhythm Patterns Using Chord Change Information, by Emiru Tsunoo, Nobutaka Ono and Shigeki Sagayama (University of Tokyo)

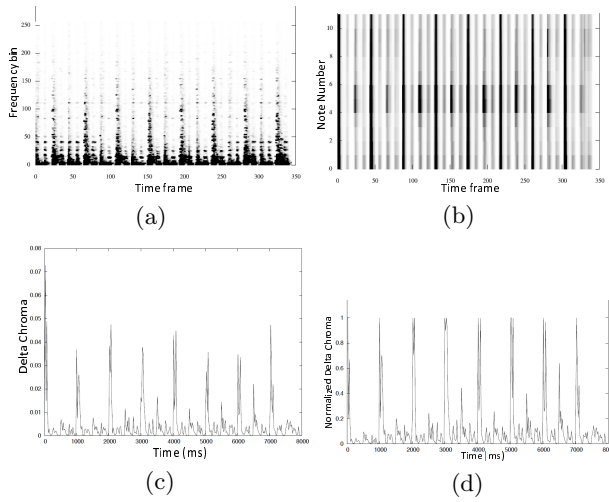


Fig. 1 (a) Spectrogram of typical music instruments, (b) its chroma vectors, (c) positive delta-chroma, and (d) normalized positive delta-chroma.

は尤度) が計算され、考えられる全ての経路の中から距離コストが最小 (尤度が最大) のものを選び出す。この小節分割のステップに和声境界情報を加えることで分割精度の向上が期待され、和声境界情報を加える考えられる方法の一つは正規化された正のデルタクロマをある重み  $w$  を掛けて加えるやり方である。ただし、和声境界情報を加えるのは one-pass DP 法でもセグメント境界 (テンプレートの頭) のみでよい。

ある時刻  $t$  での打楽器スペクトルの形状を  $r_t$  とベクトルで表し、これとあるテンプレート  $m$  の時刻 0 のスペクトル  $\mu_{m,0}$  とが距離計算される時、和声境界情報を加えた距離コスト  $c_{m,0}$  は

$$c_{m,0}(r_t) = e_{m,0,t}^T \Sigma_{m,0}^{-1} e_{m,0,t} + w \cdot d(t) \quad (2)$$

となる。ただし  $e_{m,0,t} = (\mu_{m,0} - r_t)$  で  $\Sigma_{m,0}$  はテンプレート  $m$  の時刻 0 の対角分散行列である。

#### 4 評価実験

単位打楽器パターン抽出に和声境界情報を加えたアルゴリズムを実際の楽曲に用いることによって提案アルゴリズムの有効性を評価した。用いたデータは RWC ジャンル音楽データベース [9] の中でも様々なジャンルの楽曲 8 曲 (表 1) を用いた。全ての楽曲は、サンプリング周波数 22.05kHz のモノラル信号にダウンサンプリングした。このデータセットは小節境界の正解情報を含んでいないため、正解情報は音楽経験のある男性 1 名が実際に楽曲を聞いて決定した。正のデルタクロマを計算する際は 928ms ごとに正規化を行い、距離コストに足し合わせる重みは実験的に  $w = 10.0$  と定めた。

和声境界情報を用いない場合と和声境界情報を加えた場合とで比較した結果を表 1 に示す。ここでは正解の小節境界と推定された境界線とのずれ (ms) の平均 (正は正解からの時間の遅れ、0 に近い程よい) と標準偏差を示した。これから分かるように全体的に性能が向上し、特に標準偏差の改善より平均値の改

Table 1 Comparison of bar line estimation without and with chord change information (mean differences $_{\mu}$ /standard deviations $_{\sigma}$  are shown).

Music file	Without chord change information	With chord change information
No. 6	-358.2/866.3	166.3/266.1
No. 15	310.2/429.7	216.0/349.7
No. 16	66.6/181.5	7.2/263.1
No. 19	20.5/367.8	0.1/215.6
No. 20	-168.8/312.4	-48.0/271.1
No. 21	86.6/335.7	-87.7/343.3
No. 23	-78.2/344.7	21.0/261.8
No. 26	73.4/493.1	-72.4/317.1

善の方が度合が高いことが確認できた。

#### 5 おわりに

単位打楽器パターンに基づく小節境界推定アルゴリズムに和声境界情報を加えることにより小節境界推定精度を向上させた。和声境界情報を抽出するためクロマベクトルの正の微分値を利用した。そこで得られる和声境界候補を重み付け、加えることにより既存のシステムと統合し、評価実験によって和声境界情報の小節推定における有効性を確認した。今後は、パラメータの最適化とより大規模なデータを用いた評価実験が課題と考えられる。

謝辞 本研究の一部は CrestMuse Project の支援を受けて行なわれた。

#### 参考文献

- [1] M. Goto, "A real-time music-scene-description system: Predominant-f0 estimation for detecting melody and bass lines in real-world audio signals," *Speech Communication*, vol. 43, no. 5, pp. 1783 – 1794, 2004.
- [2] M. Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," *Journal of New Music Research*, vol. 30, no. 2, pp. 159–171, June 2001.
- [3] E. Tsunoo, N. Ono, and S. Sagayama, "Rhythm map: Extraction of unit rhythmic patterns and analysis of rhythmic structure from music acoustic signals," in *Proc. of ICASSP*, 2009, pp. 185 – 188.
- [4] E. Tsunoo, G. Tzanetakis, N. Ono, and S. Sagayama, "Audio genre classification using percussive pattern clustering combined with timbral features," in *Proc. of ICME*, 2009, pp. 382 – 385.
- [5] 角尾, G. Tzanetakis, 小野, and 嵯峨山, "音楽音響信号の低音旋律パターンのクラスタリングと自動ジャンル認識への応用," in 情報処理学会研究報告, 2009.
- [6] 内山, 宮本, 西本, 小野, and 嵯峨山, "調波音・打楽器音分離手法を用いた音楽音響信号からの自動音認識," in 情報処理学会研究報告, 2008, pp. 137 – 142.
- [7] T. Fujishima, "Real-time chord recognition of musical sound: A system using common lisp music," in *Proc. of ICMC*, 1999, pp. 464 – 467.
- [8] N. Ono, K. Miyamoto, H. Kameoka, and S. Sagayama, "A real-time equalizer of harmonic and percussive components in music signals," in *Proc. of ISMIR*, 2008, pp. 139–144.
- [9] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Rwc music database: Music genre database and musical instrument sound database," in *Proc. of ISMIR*, 2003, pp. 229–230.