

音楽音響信号を対象とした和音変化時刻と和音名の同時認識

吉岡 拓也[†] 吉井 和佳[‡] 北原 鉄朗[‡] 櫻庭 洋平[‡] 尾形 哲也[‡] 奥乃 博[‡][†]京都大学工学部情報学科[‡]京都大学大学院情報学研究科知能情報学専攻

1. はじめに

音楽音響信号から和音を認識する技術は自動採譜や楽曲検索において重要である。和音認識では和音名とともに和音区間を正しく推定する必要がある。しかし、スペクトルの変化から和音の変化時刻を検出するという単純な方法は、特にドラム音を含む場合にはうまく機能しないので、和音名同定に先だてて和音区間を推定することは難しい。従来研究^{1)~3)}では、この問題を十分には解決していなかった。柏野らは和音区間を手で与えていた¹⁾。和音区間を手で与えない研究では、和音変化が起こりうる最小の時間単位として四分音符や 100msec などの短い時間を想定し、この区間ごとに和音認識を行っていた²⁾³⁾。しかし、こういった短い時間間隔では、非和声音やドラム音の存在によりスペクトルが安定しないため、正しい和音認識が困難である。

本稿では、和音区間推定を和音認識と並行して行う手法を提案する。本手法では、さまざまな和音区間の可能性を考慮しながら、和音進行に関する仮説を生成し、確信度が最大となる仮説を認識結果として出力する。この結果、和音区間の推定に和音認識結果を反映させることができるので、より正しく和音区間を求めることが期待できる。

2. 和音認識手法

本研究では、音楽音響信号から、その曲の和音進行を推定する問題を解く。和音進行とは、和音 c の時系列 $C = c_1 c_2 \dots c_m$ であり、それぞれの和音 c_i は、和音名と和音長の組、すなわち、 $c_i = (cs_i, len_i)$ で表される。

あらゆる和音長の可能性を考慮すると、処理量の爆発が生じるので、次の 2 つを仮定して各和音長の可能性を 8 通りに限定する。

- 和音の変化は、八分音符時刻上で生じる。
 - 1 つの和音の長さは、高々 1 小節 (4 拍) である。
- これらは、ポピュラー音楽において妥当な仮定である。

2.1 処理モデル

本モデルでは、和音進行仮説を追跡することによって処理が進行する。和音進行仮説とは、和音進行 C に対してその確信度 cf を付与したものを $h = (C, cf)$ である。

まず、入力音響信号から和音変化時刻の候補を抽出する。前述の仮定により、楽曲の八分音符時刻を和音変化時刻の候補とする。八分音符時刻の抽出には後藤のビート認識手法⁴⁾を用いた。処理モデルは常に一定数の確信度の高い和音進行仮説を保持しており、これらの仮説を順次更新・選択するこ

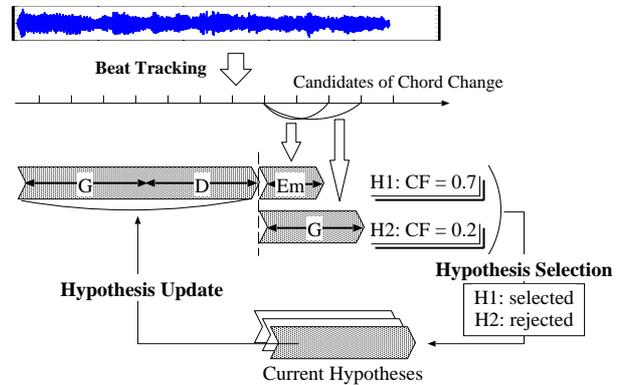


図 1 処理モデル

とによって処理が進められる。入力の終わりに到達すると、最も確信度の高い仮説の和音進行を出力する (図 1)。

2.2 和音進行仮説の生成・更新アルゴリズム

和音進行仮説は、以下に示す再帰的な枠組みにより生成・更新される。

• 初期状態

初期状態では、空の和音列 ϵ を持つ和音進行仮説 $h_0 = (\epsilon, 0)$ が 1 つだけ存在する。

• 仮説更新・選択

いま、ある和音進行仮説 $h = (c_1 \dots c_{i-1}, cf)$ が存在するとする。ここで行われる処理は、これに続く和音 c_i の候補を K 個求め (2.3 節)、それぞれの候補を仮説に付け加えたときの確信度を計算する (2.4 節) ことである。すなわち、 K 個の候補を $c_i^{(k)}$ ($k = 1, \dots, K$)、 $c_i^{(k)}$ を h に付け加えたときの確信度を $cf^{(k)}$ とすると、仮説 h は、 $h^{(k)} = (c_1 \dots c_{i-1} c_i^{(k)}, cf^{(k)})$ の K 個の仮説に更新される。

以上の処理を存在するすべての仮説に対して行った後、確信度が高い方から N 個のみの仮説を残す。現在の実装では $N = 10$ である。

• 終了条件

楽曲の終わりに到達し、仮説がそれ以上更新できなくなった時点で、仮説更新を終了し、最も確信度の高い仮説の和音進行を認識結果として出力する。

2.3 和音候補の生成

ここでは、前節の「仮説更新・選択」で行われる和音 c_i の候補 $c_i^{(k)}$ ($k = 1, \dots, K$) を求める処理について述べる。候補 $c_i^{(k)}$ とともに、その候補の信頼度も求める。

まず、前述の 2 つの仮定を満たすすべての和音長に対して以下の処理を行う。

- (1) 和音長に対応する和音区間のスペクトログラムから、各次元の値が平均律の異なる音名のパワーを表す 12 次元クロマベクトル²⁾を求め、時間方向の平均をとることで、12 次元の特徴ベクトルを得る。

Concurrent Recognition of Chord Changes and Chord Symbols for Musical Audio Signals, by Takuya Yoshioka, Kazuyoshi Yoshii, Tetsuro Kitahara, Yohei Sakuraba, Tetsuya Ogata and Hiroshi G. Okuno (Kyoto Univ.)

- (2) あらかじめ用意した各和音の学習データと、上で求めた特徴ベクトルとの距離を求める。距離尺度として、平均ベクトルとのユークリッド距離を用いた。
- (3) 距離の小さい方から L 個の和音名を求め、これを和音候補とする。また、求めた距離の逆数を総和が 1 になるように正規化したものを和音候補の信頼度とする。現在の実装では、 $L = 7$ を用いているため、 $K = 56$ となる。

2.4 和音進行仮説の確信度

本稿では、音響的特徴、和音進行パターン、ベース音に関するヒューリスティクスを用いて確信度を計算する。具体的には、和音進行 $c_1 \cdots c_m$ に対して、音響的特徴にもとづく確信度 cf_a 、和音進行パターンにもとづく確信度 cf_p 、ベース音にもとづく確信度 cf_b をそれぞれ求め、これらの積 $cf = cf_a \times cf_p \times cf_b$ を和音進行仮説 $h = (c_1 c_2 \cdots c_m, cf)$ の確信度とする。 cf_a は、前節で求めた和音候補 c_i の信頼度から求める。 cf_p は、システムの知識源に蓄積されている音楽的に妥当な和音進行パターンの適合の度合いによって求める。ポピュラー音楽では、典型的な和音進行パターンが多く用いられるので、これらに関する知識を用いて音楽的に妥当でない和音進行が出力されるのを防ぐことが期待できる。現在は、ポピュラー系の和声理論⁵⁾にもとづいて設計した $V \rightarrow VI_m$ などの 71 個の和音進行パターンを用いている。 cf_b は、

「和音 c の開始時刻では、 c の和音がベース音として鳴らされやすい」

というヒューリスティクスを仮定して、この適合の度合いによって求める。ベース音は、西洋音楽では和声と密接に関連し、和音認識の重要な手がかりとなると考えられる。

- 音響的特徴にもとづく確信度 cf_a

音響的特徴にもとづく確信度 cf_a を、和音候補 $c_i (1 \leq i \leq m)$ の信頼度の平均値で与える。

- 和音進行パターンにもとづく確信度 cf_p

$c_1 c_2 \cdots c_m$ において、知識源に存在しない和音進行の個数 n_p を求め、 $cf_p = P_p^{n_p}$ とする。このとき、和音進行パターンを適用するためには調が既知である必要があるため、並行して調を認識する。現在の実装では $P_p = 0.8$ としている。

- ベース音にもとづく確信度 cf_b

$c_1 c_2 \cdots c_m$ の各和音 $c_i (1 \leq i \leq m)$ に対して、「 c_i と同時に鳴り始める」「 c_i の和音」という 2 つの条件をみたすベース音 b を検出し、 $cf_b^{(i)} = P_b^{1-cm}$ を計算する。 cm はベース音 b の信頼度である。上記 2 つの条件をみたすベース音が存在しないときは、 $cm = 0$ として計算する。 cf_b を $cf_b^{(i)}$ の平均値として与える。現在の実装では $P_b = 0.5$ としている。ベース音の認識は櫻庭らの自動採譜システム⁶⁾を用いた。

3. 評価実験

本稿で提案した和音認識手法の有効性を調べるために、本手法と八分音符ごとの和音変化を仮定して和音認識を行う手法 (Base 手法) とを比較する。「RWC 研究用音楽データベース: ポピュラー音楽」⁷⁾ の楽曲 6 曲 (RWC-MDB-P-2001 No.14, 40, 44, 45, 46, 74) からの 1 分間の抜粋を用いて、和

表 1 和音認識率

和音進行 ベース音	Base 手法	本手法 (使用: ○, 未使用: ×)			
	—	×	○	×	○
No.14	33.3%	37.8%	53.6%	31.6%	48.5%
No.40	31.5%	50.4%	52.5%	51.7%	52.1%
No.44	25.5%	34.4%	21.5%	39.2%	52.1%
No.45	43.0%	35.8%	48.6%	43.8%	52.3%
No.46	38.5%	44.0%	42.4%	52.3%	64.9%
No.74	54.8%	50.3%	56.3%	57.9%	68.8%
平均	37.8%	42.1%	45.8%	46.1%	56.5%

音認識実験を行った。和音認識率は

$$\text{和音認識率} = \frac{\text{コード名を正しく認識した区間長の合計}}{\text{入力楽曲長}}$$

で求めた。対象とした和音の種類は major, minor, augment, diminish の triad である。

結果を表 1 に示す。本手法により、従来手法に比べて和音認識率が平均で 18.7% 向上した。また、和音進行パターンとベース音に関するヒューリスティクスを用いる場合、用いない場合に比べて 14.4% の向上が得られた。ただし、一方の手掛かりのみを用いた場合には、楽曲によっては、逆に認識率が低下した。この原因は、ベース音の音高を誤って推定したり、調を誤認識して正しくない和音名の系列を適用したためである。こうした誤りの多くは、両方の手掛かりを用いることで改善された。

本手法により和音認識の精度が改善されたが、十分な精度は達成されていない。処理精度をさらに向上させるためには、ベース音認識などの個々の処理の精度を向上し、フレーズのまとまりなどより多くの手掛かりを用いる必要がある。

4. おわりに

本稿では、実世界の音楽音響信号を対象とした和音認識を目的として、和音変化時刻と和音名を並行して認識する手法について述べた。評価実験により、本手法の有効性が確認された。今後は個々の処理を改善し、より高精度な和音認識を目指す。

謝辞 本研究の一部は、日本学術振興会科学研究費補助金基盤研究 (A) 第 15200015 号、サウンド技術振興財団研究助成、および 21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」によるものである。

参考文献

- 1) 柏野 他: “音楽情景分析の処理モデル OPTIMA における和音の認識”, 信学論, **79-D-II**, 11, pp.1847-1856, 1996.
- 2) 山田 他: “音楽音響信号を対象とした和音名同定手法: 楽曲への適用”, 音講論集, 春期, 3-7-2, pp.835-836, 2003.
- 3) A. Sheh et al: “Chord Segmentation and Recognition using EM-Trained Hidden Markov Models”, *ISMIR*, 2003.
- 4) 後藤: “音楽音響信号を対象としたリアルタイムビートトラックに関する研究”, PhD thesis, 早稲田大学理工学部, 1998.
- 5) 篠田: “新・実践コード・ワーク 1”, リットーミュージック
- 6) 櫻庭 他: “自動採譜におけるパート形成処理のための特徴量の検討”, 情処研報, 2003-MUS-51-7, pp.35-42, 2003.
- 7) 後藤 他: “RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース”, 情処研報, 2001-MUS-42-6, pp.35-42, 2001.