

ベース音高を考慮した ポピュラー音楽に対する和音進行認識

須見 康平[†] 糸山 克寿[‡] 吉井 和佳[‡] 駒谷 和範[‡] 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

音楽情報検索や推薦で重要な役割を果たし、楽曲の雰囲気や決定付ける重要な音楽的要素である和音進行の認識について、これまでに様々な研究が行われてきた。例えば、和音の特徴量としてクロマベクトルが標準的に使われており [1]、和音進行を Hidden Markov Model (HMM) でモデル化し、認識を行っている [2]。このような研究で欠如している重要な視点は、抽出対象とする和音進行を含むすべての音楽的要素が相互に関連し合っていることである。したがって、和音進行だけでなく、他の関連する要素との相関を考慮して、認識を行う必要がある。

本稿では、ベース音が構成するベースラインと和音進行との相関に着目した和音進行認識について報告する。本手法では、音高推定手法 PreFEst [3] から得られるベース音高の確率密度を利用し、和音名と和音境界を同時に求める。また、評価実験による有効性も示す。

2. 和音進行認識システム

本稿の和音進行認識システムは、ポピュラー音楽の音響信号を入力として、和音名、和音境界、調の認識を行う。なお、転調やテンポの変化はないものとする。また認識する和音の種類は表 1 の 4 種類とし、12 種類のルート音ごとに計 48 個の和音を対象とする。

和音進行認識では、和音名と和音境界の相互依存問題が存在する。これは、和音名を尤もらしくするのは正しい和音境界であり、和音境界を尤もらしくするのは正しい和音名であるというものである。このことから和音名と和音境界を同時的に認識する必要がある。

2.1 仮説探索手法

吉岡らは上記の相互依存問題に着目し、それらを同時に扱う手法として仮説探索を提案した [4]。本研究では、この手法を基にベース音高を各仮説の信頼度計算に利用する枠組みを設計する。吉岡らのシステムでは、音響の特徴量である 12 次元クロマベクトルと予め学習した単一正規分布とのマハラノビス距離、さらに頻出の和音遷移パターンやベース音に対するペナルティを設定して信頼度の計算を行っていた。

我々は、音響的特徴量の識別について Gaussian Mixture Model (GMM) を用い、また和音遷移パターンに対しては、実楽曲から学習できる和音遷移 N-gram 確率を導入することでこれらの評価値の拡張を行った。さらにベース音高と和音進行の関連を統計的に扱うことで、ベース音高の推定誤りによる認識率低下を抑制し、和音名同定、和音境界認識の重要な手掛かりとなるベース音を有効に利用することが可能となった。

信頼度の設定については 3 章で詳しく説明する。ここでは仮説の展開と探索について述べておく。

仮説の展開では、和音区間の長さを 1/2 拍から 4 拍までとし、その 8 通りの和音区間ごとに 48 個の仮説を展

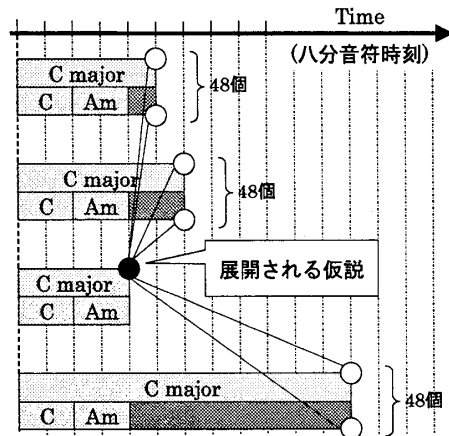


図 1: 仮説の展開

表 1: 認識対象とする和音の種類

和音名の種類	構成音
Major Triad	ルート音, 長 3 度, 完全 5 度
Minor Triad	ルート音, 短 3 度, 完全 5 度
Diminished Triad	ルート音, 短 3 度, 減 5 度
Sus4 Triad	ルート音, 完全 4 度, 完全 5 度

開する (図 1)。同和音名の連続を許容することにより、4 拍よりも長い和音に対しても認識が可能となる。

また、展開する仮説の探索手法としてビームサーチ法を用いた。ビームサーチ法では、ビーム幅を BS として、信頼度の大きい順に BS 個だけ仮説を展開し、他の仮説は今後の展開の対象としない。これは、探索空間が膨大であるため、すべての仮説を評価することによる処理量の指数的爆発を防ぐためである。

2.2 本システムにおける処理の流れ

図 2 に処理の流れを示す。ビート認識 [5] で入力楽曲の八分音符時刻を求めた後、その時刻に基づいて、ビームサーチ法により各信頼度に基づいて仮説を探索し、和音名系列、和音境界系列、調を得る。

3. 信頼度の設計

この章では仮説の評価に用いる信頼度について述べる。信頼度は以下の 3 つの評価値から決定する。

3.1 音響的特徴に基づく評価値

12 次元クロマベクトルを各フレームごとに計算し、4 種類の GMM を用いて尤度を比較することで評価を行う。

フレーム毎に得られるクロマベクトルを用いて、各和音ごとに GMM を予め学習する。ここでルート音の異なる同じ種類の和音に対して、12 次元クロマベクトルを回転させることで同一と見なし学習を行うことで GMM を 4 種に抑えることができ、学習サンプル数の増加も見込める。また識別時は、各フレームごとに得られたクロマベクトルを各ルート音に対応する幅でシフトさせることで、4 クラスの GMM に対して計 48 和音の対数尤度

Automatic Chord Recognition Based on the Pitch of Bass Sound for Popular Music: Kouhei Sumi, Katsutoshi Itoyama, Kazuyoshi Yoshii, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

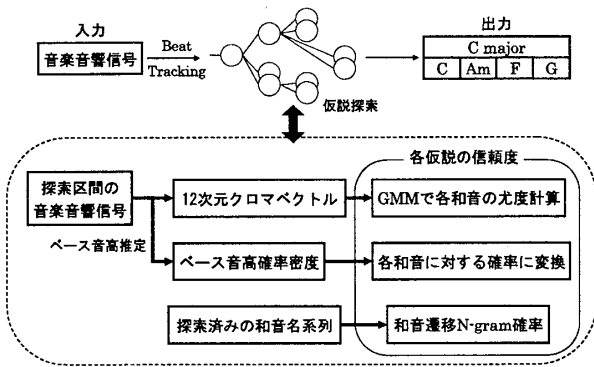


図2: 処理の流れ

を計算する。これを生成した仮説の区間長で平均したものを評価値 g_c とする。

$$g_c = \log p(X_r | \lambda_t) \quad (1)$$

ここで r はクロマの回転数 ($0, \dots, 11$), t は GMM の種類 ($0, \dots, 3$), さらに以降で現れる和音のインデックス i は $i = 4 \times r + t$ とする。

3.2 ベース音高に基づく評価値

ベース音高推定によって得られた周波数 f の重みの分布 $w(f)$ を、区間 $\tau = (\tau_s, \tau_e)$ におけるベース音の事前確率 $p(\beta_f^\tau)$ とする。

$$p(\beta_f^\tau) = w^{(\tau)}(f) = \sum_{j=\tau_s}^{\tau_e} w^{(j)}(f) \quad (2)$$

次にベース音高からのそれぞれの和音 c_i に対する確率 $p(c_i | \beta_f^\tau)$ を実楽曲から取得する。これは、正解和音ラベルとその区間におけるベース音高認識の出力結果から得られる。

この事前確率と条件付き確率をかけ合わせ、各周波数に関して足し合わせて対数スケールに変換したものをベース音高に基づく評価値 g_b とする。

$$g_b = \log \left(\sum_f p(c_i | \beta_f^\tau) p(\beta_f^\tau) \right) \quad (3)$$

3.3 和音遷移パターンに基づく評価値

本研究では、和音遷移に対する制約として和音遷移 2-gram 確率を用いた。あらかじめ長調、短調の各 2-gram 確率を実楽曲から学習する。展開した和音名を c_i , 直前の和音名を c_{i-1} として、2-gram 確率 $p(c_i | c_{i-1})$ は、それぞれの遷移が出現した回数を前の和音名の全個数で割ったものとして導出できる。この確率を対数スケールに変換して、評価値 g_p とする。

$$g_p = \log p(c_i | c_{i-1}) \quad (4)$$

4. 評価実験

提案手法の有効性を評価する実験を行った。前述の信頼度について音響的特徴、ベース音、和音遷移の組み合わせを変化させて認識率を比較した。

4.1 実験条件

評価データには、“The Beatles” の 12 枚の CD アルバムから無作為に選んだ 10 曲を用いた。これらの楽曲のステレオ音響信号の最初の 1 分を評価データとした。GMM の学習データとして、入力に用いない残りの 9 曲 (10-fold cross validation), 及び YAMAHA 社製 MIDI 音源 MU-2000 で作成した 1248 個の和音の音響信号を用いた。ここでは、音色を 6 種類、音域を 2 オクターブ、音響信号の切り出し位置を 3 種類変化させて音響信号を作成した。

表2: 10 曲の 10-fold Cross Validation 評価結果

[1]:音響的特徴, [2]:ベース音, [3]:和音遷移, [1~3]:本手法

楽曲名	[1]	[1+2]	[1+3]	[1~3]
Please Please Me	73.3	77.3	77.0	79.5
Love Me Do	68.5	78.3	72.5	78.3
Eight Days A Week	82.0	86.0	84.6	86.0
Help!	51.0	64.1	55.0	71.5
Yesterday	56.3	61.0	56.9	62.6
Eleanor Rigby	78.3	83.1	76.5	83.5
Hello Goodbye	70.6	71.2	71.6	75.3
Ob-La-Di Ob-La-Da	79.7	86.9	76.8	90.5
Let It Be	76.1	85.9	76.1	87.2
I've Got A Feeling	70.3	71.9	74.7	75.1
平均	70.6	76.6	72.2	78.9

また、ベース音高からのそれぞれの和音に対する確率の取得にも残りの 9 曲を用いた。さらに和音遷移 2-gram の学習データとして 12 アルバム中の 138 曲を長調に、13 曲を短調に使用した。

認識結果を次式に示すように、入力音響信号のうち正しく和音名を求めることができた割合により評価した。

$$\text{和音認識率} = \frac{\text{正解和音名を出力した総区間長}}{\text{入力楽曲長}} \times 100(\%)$$

学習及び評価において、Chris Harte が作成した和音名ラベルデータを正解データとして使用した。

4.2 実験結果・考察・課題

結果を表2に示す。音響的特徴のみを用いるより、ベース音、和音遷移を導入した方が認識率は上昇している。とりわけベース音と和音遷移の併用した場合、平均 78.9% という高精度の認識が可能となることが示された。

また、“Yesterday” は認識率が低くなっている。この曲では比較的珍しい遷移パターンが多用されているため、和音遷移に対する制約がうまく働かなかったためと考えられる。今後、遷移パターンの学習数を増やし、2-gram だけでなく 3-gram の導入も検討の必要がある。さらに 12 アルバム中の全 180 曲に対して実験を行う予定である。

5. おわりに

本稿では、和音進行とベースラインの相関に着目した新しい枠組による和音進行認識手法を報告した。本手法では、ベース音高推定から得られるベース音の確率密度を使用し、和音名と和音境界を同時的に認識するとともに、和音遷移 2-gram 確率を併用することによって、市販 CD の楽曲に対して平均 78.9% の認識率が達成できた。

謝辞 本研究は科研費、GCOE、CREST-Muse の支援を受けた。ご議論頂いた吉岡拓也氏 (NTT) と実験用データを提供して頂いた Chris Harte 氏 (C4DM) に感謝する。

参考文献

- [1] 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, 音響論集 (2003 春), pp.835-836, 2003.
- [2] K. Lee, M. Slaney: Automatic Chord Recognition Using an HMM with Supervised Learning, *ISMIR-2006*, pp.11-20, 2006.
- [3] M. Goto: A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals, *Speech Comm.*, 43(4):311-329, 2004.
- [4] 吉岡 他: 和音区間検出と和音名同定の相互依存性を解決する和音認識手法, 情処研報, 2004-MUS-56-6, pp.33-40, 2004.
- [5] 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キングに関する研究, 博士論文, 早稲田大学理工学部, 1998.