

ベース音高確率とクロマベクトルの相関を考慮した和音進行認識

高野 秀樹 須見 康平 糸山 克寿 駒谷 和範 尾形 哲也 奥乃 博

京都大学 大学院情報学研究所 知能情報学専攻

1. はじめに

音楽音響信号からの自動和音進行認識は音楽情報検索や推薦などの音楽情報処理システムの構築において重要な基礎技術であり、近年その研究が進められている [1, 2]. 一方、ベース音（楽曲の各時刻で鳴っている最も音高が低い音）には、以下に示す和音との関係があることが音楽的知識として知られている.

1. 和音の低音部を構成する. 特に和音の根音になる可能性が高い.
2. 和音進行を導く役割を担う.

しかしながら、和音進行認識においてベース音を用いた先行研究 [2] は少なく、その研究では和音に対してベース音高が与える影響は明示的に扱われていなかった.

本稿では、前述の知識 1 に基づき、ベース音高の推定結果から和音の根音を推定し、ベース音高とクロマベクトルとの関係からコードの種類を推定する自動和音進行認識手法について述べる. ベース音高の確率分布を PreFEst [3] によって推定し、その分布から和音の根音の尤度を計算する. クロマベクトルに基づいて根音に対する和音の種類を尤もらしさを求め、和音遷移の統計的制約を加えたうえで最尤となる和音進行を出力する. その有効性を評価するため、本手法と従来手法を用いて The Beatles の楽曲に対して和音進行認識を行った.

2. 和音進行認識処理モデル

本稿で述べる和音進行認識とは、ポピュラー音楽の音響信号を入力し、和音名の系列、和音区間の境界、調*を認識して出力する処理である. 和音の種類集合 $\mathbb{T} = \{\text{Major, Minor, Diminished, Sus4}\}$ と音名の集合 $\mathbb{R} = \{C, C\#, \dots, B\}$ を考え、認識対象とする和音の集合 \mathbb{C} を集合 \mathbb{R} (根音に相当) と集合 \mathbb{T} の直積として定義する. また、調の種類集合 $\mathbb{M} = \{\text{Major, Minor}\}$ を考え、認識対象とする調の集合 \mathbb{K} を集合 \mathbb{R} (トニックに相当) と集合 \mathbb{M} の直積として定義する.

和音名と和音境界については、楽曲中で時系列として表現される和音名系列と和音境界系列を求めることになる. すなわち、和音進行認識は入力楽曲から以下を認識する問題として定義する.

和音名系列 $C = (c_1, \dots, c_M), c_i \in \mathbb{C} = \mathbb{R} \times \mathbb{T}$

和音境界系列 $T = (t_0, \dots, t_M), t_i \in \mathbb{N}, t_0 < \dots < t_M$
 t_i は和音名 c_i と c_{i+1} との境界時刻を表す. t_0 と t_M はそれぞれ入力信号の開始時刻と終了時刻を表す. M は入力楽曲の和音区間の数となる.

Automatic Chord Recognition Considering the Relation between Bass Pitch Probability and Chroma Vector: Hideki Takano, Kouhei Sumi, Katsutoshi Itoyama, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

*楽曲中では転調しないと仮定する.

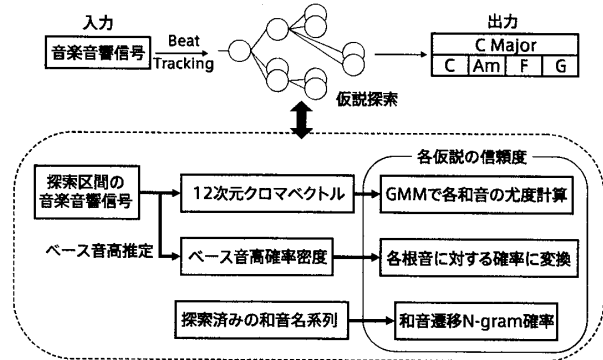


図 1: 処理の全体像.

調 $K \in \mathbb{K} = \mathbb{R} \times \mathbb{M}$

2.1 処理モデルの概要

和音進行を認識するためには和音名と和音境界を同時に認識する必要があるため、従来研究 [2] で用いられた仮説探索手法を導入する. まず、ビートトラッキング [4] により入力音響信号の八分音符区間[†]を求める. 連続するいくつかの八分音符区間で構成される和音区間に対して、ベース音高推定により得られるベース音高確率、各和音区間で展開した和音進行仮説の音響的特徴であるクロマベクトル、さらに調と探索済み和音名系列から得られる和音遷移 N-gram 確率に基づき、各和音の信頼度を計算する. ビームサーチ法により仮説を探索し、低信頼度の仮説 (和音) を枝刈りする. この操作を入力音響信号の終端時刻まで繰り返し行うことで、最も信頼度の高い和音名系列、和音境界系列、調の組を得る. 図 1 に和音進行認識の概要を示す.

3. 信頼度の設計

本章では、仮説の評価に用いる評価値を構成する 3 つの評価値について述べる. 評価値の計算には、以下の 3 つの情報を用いる.

1. 当該和音区間におけるベース音高確率 b_i
2. 当該和音区間におけるクロマベクトル \mathbf{x}_i
3. 当該和音区間以前の和音進行 c_1, \dots, c_{i-1}

i 番目の和音区間の評価値 g_i は、これらの情報が与えられた際の各和音の対数事後確率として定義される. すなわち、

$$g_i = \log p(c_i | b_i, \mathbf{x}_i, c_1, \dots, c_{i-1})$$

となる. 事後確率は、ベイズの定理を用いて以下の尤度から計算する.

$$p(c_i | b_i, \mathbf{x}_i, c_1, \dots, c_{i-1}) \propto p(b_i, \mathbf{x}_i, c_1, \dots, c_{i-1} | c_i)$$

[†]和音区間の最小単位は八分音符区間であると仮定する.

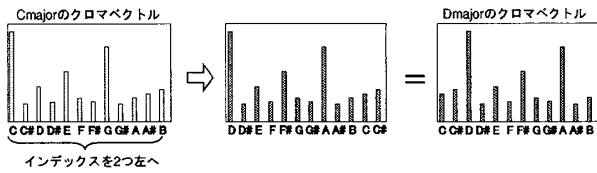


図2: クロマベクトルの回転.

第1章で述べたベース音の音楽的知識を用いて、上記の尤度が以下のように分解できると仮定する.

$$p(\mathbf{b}_i, \mathbf{x}_i, c_1, \dots, c_{i-1} | c_i) \\ = p(\mathbf{b}_i | r_i) p(\mathbf{x}_i | c_i) p(c_1, \dots, c_{i-1} | c_i)$$

ただし, $c_i = (r_i, t_i)$, $r_i \in \mathbb{R}$, $t_i \in \mathbb{T}$ である.

3.1 ベース音高確率に基づく評価値

PreFEst [3] によってベース音高確率を推定し、各根音に対応する 12 種類の混合正規分布 (Gaussian Mixture Model: GMM) を用いて尤度を計算する. 学習データ (音楽音響信号と正解和音ラベルの組) のフレームごとに得られるベース音高確率を用いて、根音ごとに GMM を学習する. 仮説探索においては、フレームごとに得られたベース音高確率に対する各根音の対数尤度を求め、その和音区間内での平均値を区間の評価値とする.

3.2 クロマベクトルに基づく評価値

12次元クロマベクトルを各フレームごとに計算し、和音の種類に対応する 4 種類の GMM を用いて尤度を計算する.

GMM の学習 学習データのフレームごとに得られるクロマベクトルを用いて、和音の種類ごとに GMM を学習する. 図2のようにクロマベクトルを「回転」させることで根音の異なる同種の和音から抽出されたクロマベクトルを同一の和音から抽出されたものと見なす. これにより、GMM の数は 4 となり、GMM あたりの学習サンプル数を増加させることができる.

GMM による尤度計算 フレームごとに得られたクロマベクトルを根音に対応する幅で「回転」させることで、それぞれの GMM から 12 の尤度を計算する. 和音区間の評価値は、和音区間内の各フレームの対数尤度の平均値とする.

3.3 和音遷移パターンに基づく評価値

本稿では、和音遷移に対する制約として和音遷移 2-gram 確率を用いる. 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. その際、和音名の 2-gram 確率を学習するのではなく、和音名を調で正規化したものの 2-gram 確率を学習する. 2-gram が出現しないパターンについては平滑化を行う. 評価値には、仮説が保有する調に基づく和音遷移 2-gram の対数尤度を用いる.

4. 評価実験

本手法の有効性を評価する実験を行った. 比較のため、従来手法 [2] を用いて同様の内容の実験を行った. 本手法がベース音高確率から $p(\mathbf{b}_i | r_i)$ を計算するのに対して、この手法は $p(\mathbf{b}_i | c_i)$ を計算している点が異なる.

表 1: 180 楽曲に対する和音進行の平均認識率.

	本手法	従来手法
和音の認識率	64.4%	65.8%
根音の認識率	69.5%	72.0%
和音の種類認識率	77.6%	77.5%

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 180 曲を用いた. これらの全曲区間の音響信号の最初の評価データとし、クロマベクトルの学習データとして入力に用いない残りの 144 曲 (5-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. また、ベース音高からのそれぞれの根音に対する確率の取得にも残りの 144 曲を用いた. さらに和音遷移 2-gram の学習データとして 12 アルバム中の変調する曲、調のない曲を除いた 151 曲中の 138 曲を長調に、13 曲を短調に使用した. 学習及び評価においては、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. 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.

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

4.2 結果と考察

表 1 に、180 曲に対する和音進行認識結果の平均正解率を示す. 従来手法と、その簡略化に相当する本手法との認識率がほぼ等しいことから、従来手法には冗長な計算が含まれていたこと、本手法ではその削減ができたことが示唆される.

5. おわりに

本稿では、ベース音高と音響特徴による自動和音進行認識について述べた. 実験結果より、従来手法ではベース音高の利用方法についてデータによっては冗長な計算となっていたことが示唆された. 今後は、最適な計算の自動選択手法の構築やこの冗長性の削減による計算時間の評価などを行う予定である.

謝辞 本研究の一部は、科研費、グローバル COE、Crest-Muse の支援を受けた. コードラベルの Ground Truth は Chris Harte 氏 (C4DM) から提供いただいた.

参考文献

- [1] C. Harte *et al.*, “Automatic Chord Identification using a Quantised Chromagram”, Proc. AES Convention, 2005.
- [2] 須見他, “ベース音高を考慮した音楽音響信号に対する和音進行認識”, 情処全大 2X-5, 2008.
- [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 Communication, Vol. 43, No. 4, pp. 311–329, 2004.
- [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キングに関する研究”, 博士論文, 早稲田大学理工学部.