

調波音・打楽器音分離手法を用いた音楽音響信号の調と和音の同時認識*

内山裕貴, 宮本賢一, 西本卓也, 小野順貴, 嵯峨山茂樹 (東大院情報理工)

1 はじめに

本稿では、音楽情報検索や自動採譜等への応用を目的とした、音楽音響信号から調と和声進行の認識問題を扱う。我々は、打楽器等の非調波的な音を含む楽曲に対し、宮本らの調波音・打楽器音分離手法 [1] により得られた調波音から求めたクロマを特徴量に用いて和音を認識する手法を提案した [2]。

西洋音楽をはじめとした調性音楽においては、調や和声が音楽の構造を決める基本的な要素であり、和声と調を認識したい。和音に加えて調が分かることで、和音の機能が明確になり、音楽情報処理での応用や、和音認識率の向上が期待できる。

調を認識するとき、使用されている和音が大きな手がかりになるが、逆に調がその和音の手がかりになる。そこで調と和音を同時に認識する手法を提案する。

2 調と和音の同時認識

2.1 HMM による調と和音のモデル化

和音を認識するとき問題となるのは、和音がそのまま演奏されるわけではなく、和声内音の省略や非和声音の挿入があるため、フレームごとみで認識するのは難しいということである。そこで、HMM による定式化を用いて和音認識を行った [2]。和声進行を和音のバイグラムでモデル化し、その遷移確率によって和音間の遷移のしやすさ偏りを利用できる。また、出力確率によって、各和音からどの音が演奏されやすいかを反映することができる。

先行研究 [2] の調を考慮しないモデルでは、和音間の遷移確率は二つの和音だけで決まっていたが、同一和音間でも調によって機能が異なり、遷移確率が異なる。そこで、相互依存の関係にある調と和音を同時に推定するため、Fig. 1 のような和声モデルを提案する。このモデルは吉川 [3] らの旋律への自動和声付けで用いられた調と機能音声モデルと似ているが、ポピュラー音楽では、機能音声で扱うことが難しい和音があるため、我々は調と和音の組み合わせのモデルを用いる。平行長短調を同一調と扱い 12 種類の調毎に、和音がエルゴディックに遷移するモデルを作成す

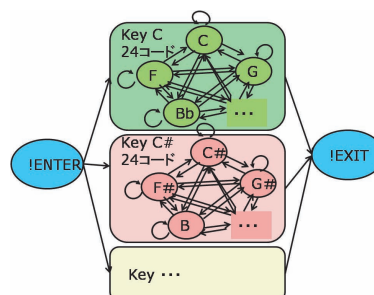


Fig. 1 和声モデル

る。なお、転調のある曲については、吉川らのように調の間の遷移確率を定めることによって対応できると思われるが、本稿では扱わない。特徴量に関しては、先行研究と同様に、調波音を強調したクロマベクトルを用いた。

2.2 パラメータ学習方法

学習は、すべての音楽音響信号と和音ラベルを八長調に移調して行い、得られた八長調の遷移確率と出力確率を、根音をずらして各調のパラメータとした。これにより、学習に必要なデータを増やすことなく、状態を増やす事が出来る。

なお、調を考慮するとき、固有和音は出現回数が多く十分学習できるが、非固有和音は出現回数が少ないため、学習データを増やしても、十分学習することが困難である。学習データの不足は、文法モデルに関しては、backoff 平滑化を用いた。音響モデルに関しては、同種の和音を根音が同一になる様に移調して平均したものを、もとの和音に移調するして利用した。

2.3 アルゴリズム概要

アルゴリズムの概要は以下のようになる。

- (1) 入力信号に対し、宮本ら [1] の調波音・打楽器音の分離を用い、調波音を得る。
- (2) スペクトログラムをオクターブ間で足し合わせて、クロマベクトルを計算する。
- (3) クロマベクトルと和音ラベルデータをすべて八長調に移調し、出力確率と遷移確率の学習を行い、根音をずらして、各調の出力確率と遷移確率を求める。
- (4) 原調のクロマベクトルを用いて、調と和音を認識する。

*"Automatic Key and Chord Detection Using Harmonic Sound Emphasized Chroma from Musical Acoustic Signal," by Yuki Uchiyama, Ken-ichi Miyamoto, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama, Graduate School of Information Science and Technology, The University of Tokyo.

条件	混合信号	調波信号
(a)	70.4%	76.4%
(b)	72.4%	76.5%
(d)	73.8%	77.6%
(d)	74.7%	78.3%
(e)	72.2%(19.8%)	77.8% (17.2%)

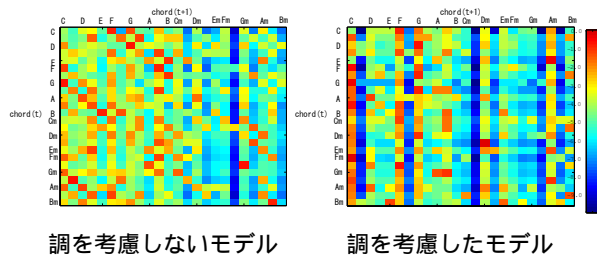


Fig. 2 遷移確率 $\log(P(c(t+1)|c(t)))$

3 評価実験

3.1 実験条件

2章で提案したアルゴリズムを評価するための実験を行った。学習及び認識には The Beatles の 12 枚のアルバムに含まれる 180 曲の楽曲中、転調のない 162 曲を用いた。学習・正解ラベルは C. Harte[4] らの作成したものを利用した。和音の種類は、長三和音と短三和音の 24 種類としたが、このラベルは、長三和音、短三和音以外の和音も含むため、根音と第三音から、長三和音または短三和音とみなした。調については人手で判定し与えた。学習はデータセットを 2 分し、学習データと認識データを入れ替えて交互に実験した。

また、比較のため、以下の 5 条件で実験した。

- 従来の調を考慮しないモデル。音響モデルの平均化なし。
- 調を考慮したモデル。ただし、調を既知として、すべてのクロマベクトルを八長調に移調して認識。音響モデルの平均化なし。
- 調を考慮しないモデル。
- 調を考慮したモデル。ただし、調を既知として、すべてのクロマベクトルを八長調に移調して認識。
- 本校で提案した調を考慮したモデルにより、和音と調を同時に推定する。

3.2 結果と考察

学習した遷移確率を Fig. 2 に示す。調を考慮しないモデルは、遷移確率は相対度数に大きく依存し、4 度や 5 度の斜め方向に確率が高い部分が見られる。そのため、調が八長調であるが、和音 F から和音 B \flat などの非固有和音への遷移確率も高い。一方、調を考慮

したモデルでは、このような非固有和音への遷移確率は低く、固有和音、特に主要三和音に遷移する和音が多い。

評価実験の結果を Table 1 に示す。調を考慮しないモデルの (a) に比べ、調を考慮した (b) は全体の認識率が向上している。一部の曲では認識率が下がっており、例えば、“Come Together” では認識率が、56.3%から 24.3%に低下した。この曲は二長調であるが、非固有和音の Dmin 和音があり、調を考慮することで、学習データの不足のため認識が難しくなったのだと考えられる。そのため (d) のように音響モデルの平均化を行うことによって、認識率が向上した。ただし (c) に関しても、状態ごとの学習データが増え、認識率が向上した。

調を同時に推定した (e) では、(d) に比べて認識率が低下した。なお、調は間違えているものの和音名自体は正しく認識したものが多かった。これは、遷移確率が似ている調があるためである。和音名が同一の和音は調に依らず同一の音響モデルを用いるため、和音名は正しく認識できるが、調を認識するのが困難であった。

4 おわりに

本稿では、調を考慮した HMM と調波音を強調したクロマベクトルを用いて、調と和音を同時認識する手法を提案した。調が既知の場合には認識率が向上し、調を用いる有用性が確認できたが、調未知の場合には調の認識が難しく認識率が向上しなかった。今後は、調の認識率を向上させるため、調ごとに異なる音響モデルを用いる手法を検討する予定である。

謝辞 本研究の一部は、科学技術振興機構 CREST の補助を受けて行われた。

参考文献

- 宮本他, “スペクトログラムの滑らかさの異方性に基づいた調波音・打楽器音の分離,” 日本音響学会春季研究発表会講演論文集, 2008.
- 内山他, “調波音を強調したクロマに基づく音楽音響信号からの自動和音認識,” 日本音響学会春季研究発表会講演論文集, pp. 901-902, 2008.
- 吉川他, “HMM を用いた旋律への自動和声付けと調性推定,” 日本音響学会秋季研究発表会講演論文集, pp. 533-534, 2000.
- C. Harte *et al.*, “Symbolic representation of musical chords: a proposed syntax for text annotations,” Proc. ISMIR, pp. 66-71, 2005.