3Q-02

ビート位置依存隠れセミマルコフモデルに基づく音楽音響信号に対するコード認識

丸尾 智志[1]　前澤 陽[2]　中村 栄太[1]　糸山 克寿[1]　吉井 和佳[1]

[1]京都大学 大学院情報学研究科 知能情報学専攻　[2]ヤマハ株式会社

1. はじめに

音楽音響信号に対する自動コード認識は，音楽情報処理の分野における基本的な課題の一つである．楽曲のコードパターンは楽器演奏の手助けとなるだけでなく，ジャンル分類 [1] や楽曲推薦 [2] などにも利用できることがその理由である．

コード認識の従来法は一般的に音響特徴量ベクトルの抽出とその分類で構成される．最もよく用いられている音響特徴量は 12 のピッチクラス（C，C♯，…，B）のエネルギーの分布を表現した 12 次元クロマベクトル [3] である．このクロマベクトルをフレームごとやビートごとといった短い区間ごとに計算することで特徴量の抽出が行われる．一方で特徴量の分類には，コードの遷移確率とコードの種類ごとのクロマベクトルの出力確率を表した隠れマルコフモデル（HMM）がよく用いられる [4]．コード認識の精度改善のために，コードの特徴をよく表した様々な特徴量の抽出方法が提案されている [5,6]．それに比べて，HMM を用いた特徴量分類の従来法はコード遷移の特徴が十分に反映されておらず，スムージングとしての役割が大きくなっている [6]．コード遷移の特徴として、遷移のタイミングにビート位置に連動した一定の傾向（例えば，1 拍目や 3 拍目などで遷移しやすい）が見られることが挙げられる．また，同じコードが続く長さにも一定の傾向（例えば，1 小節間やその半分の長さなどキリのよい長さ継続する）が見られる．これらの特徴を考慮したモデルを使用することで，より正確なコード認識を行うことができると考えられる．

本稿では，コード遷移とコード継続長のビート位置依存性を考慮したモデルによるコード認識手法を提案する．提案手法では，コード遷移の特徴を反映するために，隠れセミマルコフモデル（HSMM）によってコードの遷移をモデル化する．これにより，コードの継続長の偏りをモデル化し，同じコードが数ビート継続する確率が指数減衰してしまうという HMM によるコード認識の問題を解消する．また，ビート位置ごとに異なる遷移確率，継続長確率を使用することで，コード遷移のタイミングのビート位置依存性を反映する．

2. 提案手法

ここでは，提案手法であるビート位置依存 HSMM に基づくコード認識 (図 1) について述べる．提案手法では，まず音楽音響信号から [7] で提案した手法により一般クロマベクトルと低音クロマベクトルを抽出する．続いて，抽出したクロマベクトルを用いて，ビート位置依存 HSMM によるコード認識を行う．

2.1 問題設定

コード認識は，音楽音響信号からコードラベルの系列を得る問題である．すなわち，対象の信号から特徴量を抽出し，識別器を用いてそれに対応するコードラベルの系列を得ることが目標となる．

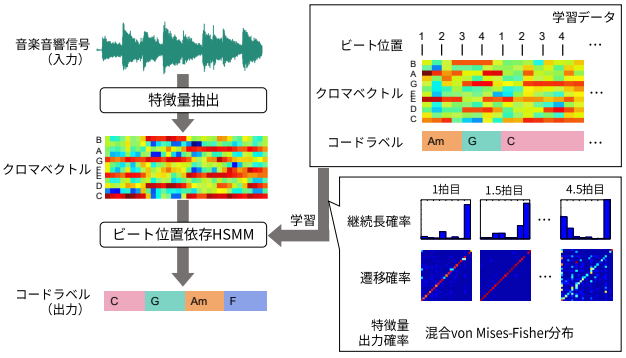

図 1: 提案手法の全体図

本稿では，以下の条件下でコード認識を行う．

- ビートおよび小節線の位置は [8] を用いてあらかじめ推定する．
- コードの境界はハーフビートの位置に存在しているとする．
- 認識するコードの種類は各ルート音に対して “maj”，“min” の 2 種類（MajMin），またはそれに “7”，“maj7”，“min7” を加えた 5 種類（Seventh）とし，どちらの場合も “no chord” を含む．他のコード（“aug” や “dim” など）は [5] を参考にいずれかに分類する．

2.2 クロマベクトルの抽出

コード認識の特徴量には，[7] で使用した非負値行列因子分解（NMF）によって抽出されるクロマベクトルを使用する．基底スペクトルとして MIDI ノートナンバー 21〜108 に対応する 88 個の音高のスペクトルを用意し，対象の音楽音響信号のスペクトログラムに対して NMF を行うことで，各音高の音量変化に対応するアクティベーションを得る．続いて，得られたアクティベーションから以下の式によりフレーム t におけるクロマベクトルの各次元の値 x_{nt} を計算する．

$$x_{nt} = \sum_{p:p\equiv n(\mathrm{mod}\ 12)} \exp\left\{-\frac{(p-\mu_p)^2}{\sigma^2}\right\} \cdot h_{pt} \qquad (1)$$

ここで，n は 12 のピッチクラス，h_{pt} はフレーム t における MIDI ノートナンバー p の音高のアクティベーション，$\exp\{\cdot\}$ は音高ごとの重みである．一般クロマベクトルを計算する際は，$\mu_p = 65$，$\sigma = \frac{44}{3}$，低音クロマベクトルを計算する際は，$\mu_p = 43$，$\sigma = \frac{22}{3}$ とする．最後に，フレーム単位で得られたクロマベクトルに対してハーフビートの区間ごとに平均を取ることで，各ハーフビート位置でのクロマベクトルを計算する．各クロマベクトルは平均が 0，L2 ノルムが 1 となるように正規化する．

2.3 ビート位置依存 HSMM に基づくコード認識

ビート位置依存 HSMM は，状態をコードと調の対とすることで楽曲中での転調にも対応することができるよ

Automatic Chord Recognition based on a Beat-Position-Dependent Hidden Semi-Markov Model: *Satoshi Maruo (Kyoto Univ.), Akira Maezawa (Yamaha Corp.), Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii (Kyoto Univ.)*

うにした転調 HMM [7] を HSMM に拡張し，さらにビート位置ごとに異なる遷移確率と継続時間長を与えたものである．HSMM の HMM との大きな違いは，継続長確率の存在である．HMM では同一コードが数ビート継続する場合は自己遷移の連続により確率が指数減衰してしまう．HSMM では同一コードが継続する確率は継続長確率によって表現されるため，この問題を解消することができる．また，継続長確率を使用することにより，コードの継続長の偏りをモデル化することができる．さらに，ビート位置依存 HSMM は，ビート位置によって異なる継続長確率および遷移確率をもつ．これにより，コード遷移が特定のビート位置で起こりやすいという特徴を反映することができる．

ビート位置依存 HSMM の学習は，以下の 3 つの確率を計算することで行われる．

継続長確率　各ビート位置からの同一コードの継続ビート数を数えることで，全てのコードで共通のビート位置ごとの継続長確率を学習する．

遷移確率　各ビート位置でのそれぞれの遷移の発生回数を数えることで，ビート位置ごとのコードと調の遷移確率を学習する．遷移前の調の主音を C に移調し，それに伴って遷移後の調の主音および遷移前後のコードのルート音もシフトさせる．したがって，遷移前の調の主音が C である遷移のみ遷移確率を学習する．他の 22 の調における遷移確率は要素を並び替えることで得られる．HSMM では一般的に自己遷移は認めないが，本稿では自己遷移もゆるものとして学習する．

特徴量出力確率　各コード区間におけるクロマベクトルの出力確率を混合 von Mises-Fisher 分布で学習する．各コードのルート音はクロマベクトルの要素を巡回シフトすることで C に統一する．したがって，コードタイプごとに出力確率を学習することになる．他のルート音の出力確率はパラメータを並び替えることで得られる．

以上により学習された HSMM に対して，ビタビアルゴリズムによる最尤経路探索を行うことで，コード認識を行う．

3. 評価実験

提案手法の精度を評価するために，実録音の楽曲のコード認識を行った．

3.1 実験条件

コード認識の評価には The Beatles データセット [1] の 179 曲を使用した．すべての楽曲のサンプリングレートは 16 kHz である．各音高の音量推定を行う際のスペクトログラムは，周波数の間隔は 25 cent，ステップ幅は 25 ms の定 Q 変換により計算した．HSMM の状態継続長の最大値は 8 ハーフビートとした．実験は 10-cross validation により行い，認識率は楽曲の長さに対してコード認識結果が正解であった時間の割合として計算した．比較として，提案手法（BD-HSMM）の他に，HMM，HSMM，ビート位置ごとに異なる遷移確率を持つビート位置依存 HMM（BD-HMM) によるコード認識を行った．

3.2 実験結果

表 1 に実験結果を示す．MajMin の認識の場合，ビート位置依存 HMM を用いることで HMM を用いる場合よりも認識率が 1.2 ポイント向上した．また，ビート位置依存 HSMM を用いることで，HSMM を用いる場合よりも認識率が 1.3 ポイント向上した．このことから，ビート位置ごとに異なる遷移確率を使用することはコード認識において有効であることが示唆される．

表 1: 実験結果

	HMM	HSMM	BD-HMM	BD-HSMM
MajMin	80.7	81.0	81.9	82.3
Seventh	70.1	70.1	70.9	71.0

一方で，HSMM を用いた場合の認識率は，HMM を用いた場合の認識率に比べて 0.3 ポイント向上，ビート位置依存の場合は 0.4 ポイント向上とわずかな改善しか得られなかった．これは，HSMM の状態継続長確率をすべての状態で同一にしたことが一因であると考えられる．コードの役割によって継続ビート数の分布には違いがあると考えられるため，ビート位置だけでなくコードの種類でも継続長確率を使い分けることでより大きな精度の向上が得られる可能性がある．

また，Seventh の認識の場合も MajMin の認識と同様の傾向が見られた．ただし，認識率向上の幅は MajMin に比べると小さかった．これは，認識するコードの種類数が増えたことにより，遷移確率が十分に学習できなかったことが原因であると考えられる．

4. おわりに

本稿では，コード遷移の特徴を反映したビート位置依存 HSMM に基づくコード認識手法を提案した．実験により，提案手法によりコード認識率が向上することが示された．しかし，認識するコード数が多い場合は提案手法による認識率の改善があまり見られなかった．今後より複雑なコードを認識するためには，学習データの充実および効率的な学習が必要となる．また，本稿ではあらかじめ推定したビートを使用してコードの認識を行ったが，ビートに誤りがある場合はコード認識精度が悪くなると考えられる．そのため，コードとビートの相互依存性を考慮して，それらを同時に認識するようなモデルを用いることで，より正確なコード認識を行うことができると考えられる．

謝辞　本研究の一部は、JSPS 科研費 24220006, 26700020, 26280089，JST CREST プロジェクトの支援を受けた．

参考文献

[1] C. Pérez-Sancho and *et al.* Genre classification using chords and stochastic language models. *Connection science*, 21(2-3):145–159, 2009.

[2] M. Goto and *et al.* Songle: A web service for active music listening improved by user contributions. *ISMIR 2011*, 311–316, 2011.

[3] T. Fujishima. Realtime chord recognition of musical sound: A system using common lisp music. *ICMC 1999*, 464–467, 1999.

[4] K. Lee and M. Slaney. Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio. *IEEE TASLP*, 16(2):291–301, 2008.

[5] M. Mauch. *Automatic chord transcription from audio using computational models of musical context.* PhD thesis, School of Electronic Engineering and Computer Science Queen Mary, University of London, 2010.

[6] T. Cho and J. P. Bello. On the relative importance of individual components of chord recognition systems. *IEEE TASLP*, 22(2):477–492, 2014.

[7] 丸尾智志 他. 音楽音響信号に対する歌声・伴奏音・打楽器音分離に基づくコード認識. **第** *108* **回音楽情報科学研究会**, 2015.

[8] S. Durand and *et al.* Downbeat tracking with multiple features and deep neural networks. *ICASSP 2015*.

[1] http://www.isophonics.net/datasets