

E-029

状態共有型 HMM に基づく音楽音響信号からの自動和音認識の検討

A Study on Automatic Chord Recognition
Using Shared-States HMM from Musical Acoustic Signals伊藤 綾†
Aya Ito酒向 慎司†
Shinji Sako北村 正†
Tadashi Kitamura

1. はじめに

計算機やネットワークの普及により、音楽コンテンツの多様化・大規模化が進むなかで、より柔軟な音楽検索やコンテンツ作成支援のための環境はまだ十分ではなく、音響信号の内容そのものを音楽的な側面で解析する技術が求められている。本稿では音楽の内容を表す情報として和音進行に注目し、音響信号からの自動和音認識を扱う。これは一般に和声解析や自動採譜に寄与する技術である一方で、音楽検索や音楽分類への応用も期待できる。

自動和音認識ではクロマベクトルを特徴量とした隠れマルコフモデル (HMM)[1] がよく用いられるが、本稿では前後関係を考慮した環境依存 HMM を使い、クラスタリングによるモデルパラメータの共有の有効性を検討する。

2. 状態共有型 HMM による和音認識

2.1 特徴量:クロマベクトル

和音には複数の転回形が存在し、音高の配置が異なる場合でも、構成音が同一であれば同一の和音として認識される。そこで、パワースペクトルを音名ごとにオクターブ間で足し合わせた特徴量が有効と考えられる。この特徴量はクロマベクトルと呼ばれ、音響信号からの和音認識でよく利用されている。本稿では1つの半音が1つの次元に対応する12次元のクロマベクトルを用いる。図1にクロマベクトルの概念を示す。このクロマの時系列ベクトルを演奏パターンと見なす。

2.2 HMM による和音進行のモデル化

1つの和音から生成されるクロマにはある程度の揺らぎがあり、これは出力確率分布で扱うことができる。また和音の遷移についても和声学に基づく一定の規則があり、多くの事例から学習することで確率的に扱うことができる。これらを考慮したモデルとして、自動和音認識には HMM がよく用いられる。調性音楽では一般的に、和音から楽曲が作成されると考えることができるため、和音を隠れた状態系列とし、演奏パターンは各和音の出力確率分布から生成されるとみなす。

本研究では、1つの和音が1つの状態に対応し、全ての和音へ遷移可能な ergodic HMM で和音進行をモデル化する(図2)。和音間の遷移のしやすさは調性に大きく影響されるため、本来は調ごとにモデルを作ることが望ましいが、ここでは全ての調を1つのモデルで扱う。

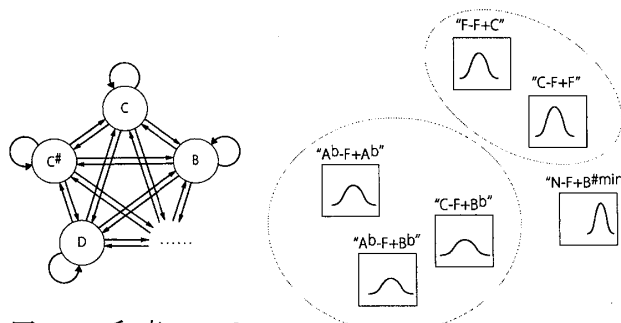


図2: 和音 ergodic HMM

図3: モデルパラメータのクラスタリング

2.3 環境依存 HMM の利用と状態共有

HMM を用いた和音認識は過去にも研究例があり、山らは、クロマベクトルはその時刻の和音のみによって確率的に生成されると仮定し、各和音のクロマから出力確率分布を学習した[2]。一方 H. Papadopoulos らは、出力確率分布を音響信号から学習せず、音楽理論に基づいて定めた[3]。出力確率分布は単一の正規分布を仮定し、倍音の影響を考慮することで従来の手法よりも高い認識率を得た。

実際の演奏では和声内音が省略されることも多く、1つの和音から生成されるクロマのパターンは多様であると考えられる。そこで、ある和音から生成されるクロマベクトルは、その前後の和音進行にも影響されると仮定し、音声認識の分野でよく用いられる triphone モデルを和音認識に適用する。

ただし、あらゆる trigram 和音進行パターンについて十分な学習データを得ることは難しいため、過学習の問題が予想される。これを避けるため、HMM の状態共有を行う。図3のように、中心和音が同一の tri-gram モデルの中でモデルパラメータが近いものをクラスタリングすることで頑健なモデルを作成する。

3. 実験

3.1 実験条件

本システムを The Beatles のアルバム “With the Beatles” を用いて評価する。楽曲は CD のデータをモノラル化し、11025Hz にダウンサンプリングして用いた。時間分解能は 0.1s とし、55.0Hz(A0) から 3729.3Hz(Bb6) までの6オクターブの中心周波数をもつ定 Q フィルタバンクを用いて時間周波数解析を行った。

和音の語彙は major, minor の 24 種類とし、それ以

†名古屋工業大学, Nagoya Institute of Technology

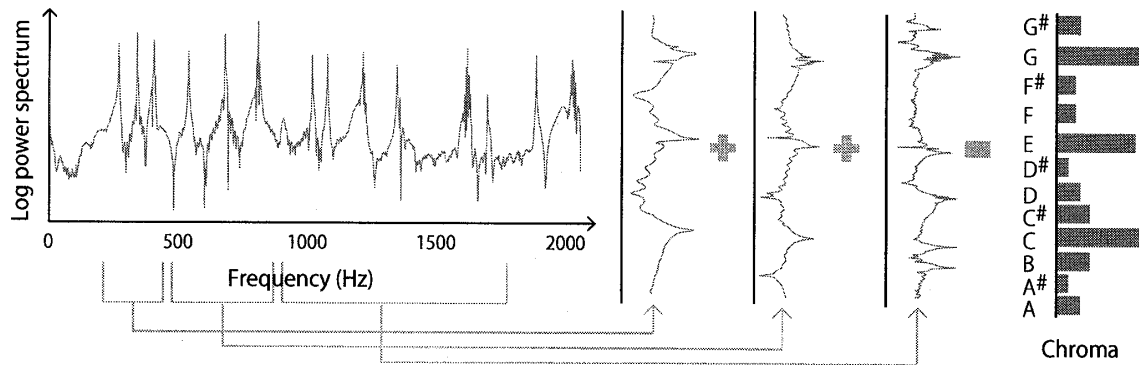


図 1: クロマベクトルの概念

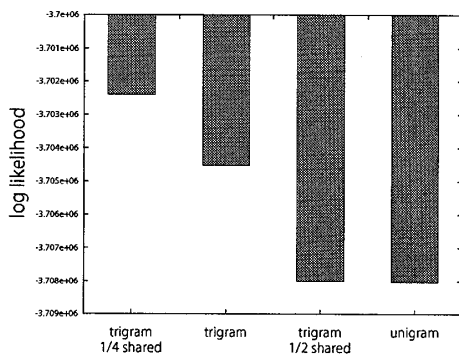


図 4: 各モデルの Viterbi パスの尤度

外の和音は第3音に着目して major, minor に近似した。各和音のクロマベクトルの出力確率は単一の多次元正規分布と定め、アルバム内の14曲中13曲から各モデルの平均と分散共分散行列を学習した。unigram, trigram, 状態共有型 trigram の性能を比較する。状態共有型 trigram は和音毎に共有構造を行い、各和音のモデル数が1/2, 1/4になるものを作成した。

3.2 モデルの評価

クラスタリングによる共有の効果を確認するため未学習データに対する Viterbi パスの尤度によってモデルを評価する。先の実験条件の下で Viterbi アライメントをとった際の尤度を図4示す。

モデル数を1/4まで共有した trigram で尤度が最も高くなっており、unigram では一番低くなっている。unigram での尤度が低いのは、和音から生成されるクロマを単一の出力確率分布でモデル化することで、様々なパターンが平滑化されているためと考えられる。また、trigram モデルを共有することで尤度が向上している。これは、1つの和音を前後関係によって細かく分類するという手法では、学習データで出現頻度の低いものは過学習が起り、ある特定の楽曲の特定の和音のみを表すようなモデルになるためである。そこで、細分化されたモデルをもう一度類似した和音と共有してモデルを構築す

ることで、より高い尤度を得ることができた。

4. むすび

本稿では、音楽音響信号からの自動和音認識に状態共有型 HMM を適用することを提案した。観測される演奏パターンがその時刻の前後の和音からも影響を受けることに着目し、モデルを trigram へ拡張することで性能の向上を図った。ここで、モデルを細分化することによる過学習が和音認識性能を低下させる可能性があるため、類似したモデルパラメータをもつ状態を共有することでより頑健なモデルを作成した。実験から、状態共有型 trigram モデルを用いることで Viterbi パスの尤度が向上することが確認でき、音響モデルを細分化することで和音認識性能を向上させるという期待が示された。

今回は音響モデルのみの評価を行ったが、和音は和声学に従って進行するため、これに相当する言語モデルを併用して和音認識をすることが課題である。

また、共有構造をモデルの分布間距離のみで決めたと、和音の前後関係等の音楽的知見に基づいてモデルを共有することが課題として挙げられる。さらにクロマベクトルの分布は和音の性質に由来しているのみでなく、楽曲のジャンルや構成楽器等にも依存していると考えられるため、楽曲自体もある分類基準で区別することが有効と考えられる。

謝辞 本研究の一部は、(財)電気通信普及財団 平成21年度研究調査助成による支援を受けた。

参考文献

- [1] A. Sheh, et al., "Chord segmentation and recognition using EM-trained hidden Markov models", Proc. ISMIR, pp. 183–189, 2003.
- [2] 内山 裕貴, 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集, pp. 901–902, 2008.
- [3] H. Papadopoulos, et al., "Large-scale study of chord estimation algorithms based on chroma representation and HMM", Proc. CBMI, pp. 53–60, 2007.