Vol.2010-MUS-88 No.4
2010/12/4

パラメータ共有 HMM に基づく 音響信号からの自動和音認識の検討

伊藤 綾[†1] 酒向慎司[†1] 北村 正[†1]

本稿では，隠れマルコフモデルに基づいた自動和音認識において，和音に依存した音響特徴を精密にモデル化する試みについて報告する．観測されるクロマベクトルの系列が，該当の和音だけでなく，一連の和音連鎖に依存していると考え，単独の和音だけでなく，前後の和音に依存した詳細な環境依存の和音連鎖 HMM を考える．このようなモデルの詳細化によって，統計モデルの学習が困難になるため，クラスタリングによるモデルパラメータの共有化を行う手法を提案し，その有効性を検討する．

A Study on Automatic Chord Recognitiono Using Tied Parameters HMM from Musical Acoustoic Signal

Aya Ito,[†1] Shinji Sako[†1] and Tadashi Kitamura[†1]

In this paper, we propose a technique for detail acoustic modeling of HMM-based auto chord recognition from music signals. In our approach, we use a context dependent chord unit that models dependently the previous and successive part of a chord. However, this strategy often result in over-fitting because of short of the training data. To avoid such problem, we use decision-tree state clustering technique that aims to tying parameter of HMM.

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・大規模化が進んでいる．例えば，iPod などに代表される音楽プレーヤーの小型化や大容量化によって，多量の楽曲を手軽に扱えるようになった．その一方で，多量の音楽をより柔軟で効率的に検索するための技術が求められている．音楽の分類には，歌手や曲名などの形式的な情報を利用することが多いが，ジャンルなどの分類では一意に定まらないなどの問題もある．一方で，協調フィルタリングを利用したアプローチでは，利用者の行動履歴をもとに楽曲を推薦する手法が，音楽配信サービスなどで採用されている．これらの方法に共通する点として，コンテンツの内容を検索に直接的に利用していないことが挙げられる．このような問題意識から，楽曲そのものの情報を音楽検索に利用するために，音楽的な側面からコンテンツを解析する技術が求められており，これまでにも様々な研究が展開されている．

我々は，音楽の内容を表す上で，有効な特徴の一つである和音進行に注目し，和音進行によって楽曲の類似性を分類する試みに取り組んできた[1)]．本稿では，このような類似曲分類への応用を念頭に，音響信号からの自動和音認識手法として，従来から提案されている隠れマルコフモデル (Hidden Markov Model; HMM) による手法において，和音に依存した音響特徴を精密にモデル化する手法について検討し，単独の和音ではなく，前後の和音に依存した詳細な環境依存の和音 HMM を考える．このようなモデルの詳細化によって，統計モデルの学習が困難になるため，クラスタリングによるモデルパラメータの共有化を行う手法を提案し，その有効性を検討する．

2. HMM による和音認識

音響信号からの自動和音認識は，和声解析や自動採譜のための主要な技術となる一方で，音楽検索や音楽分類といった分野など広く応用が期待されている．前節で述べたように，自動和音認識の研究では，和音の音楽理論的な制約と音響的特徴を HMM の枠組みで扱ったものがいくつか提案されている．これは，和音ごとに発音される音の偏りがあることから，ある短時間スペクトルから抽出される音名に特化した特徴（クロマベクトル）を観測系列とし，和音の進行に見られる一定の規則や統計的な性質を利用した制約により，楽曲の連続した音響信号から和音の進行を求めるものである[3)]．

2.1 クロマベクトルによる音響特抽出

和音には複数の転回形が存在するため，音高の配置が異なる場合でも，構成音が同じであれば同一の和音として認識される．これを音響信号に含まれる音名の特徴抽出に利用すると，短時間パワースペクトルをオクターブ毎に帯域を分割し，オクターブ間で同一の音名を足し合わせた特徴量が有効であると考えられる．このような特徴量はクロマベクトル (Croma Vector) と呼ばれ，音響信号からの和音認識のほか，音楽の特徴抽出としての有効

†1 名古屋工業大学大学院工学研究科
Graduate School of Engineering Nagoya Institute of Technology

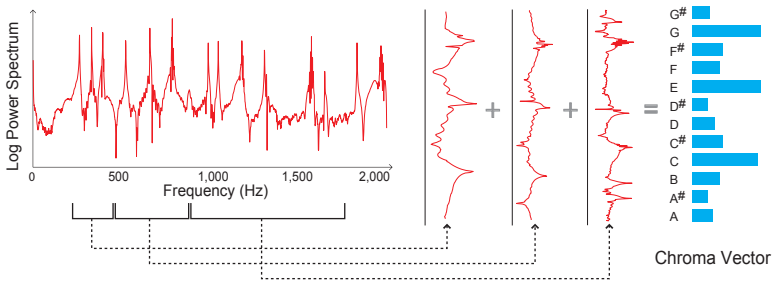

図 1 クロマベクトルの概要
Fig. 1 A summary of chroma vector

性が示されている[2].

クロマベクトルの作成法はいくつか提案されているが，本稿では式 1 のように 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．ただし，$H(f,t)$ はスペクトログラムにおける周波数 f，時刻 t でのパワーを表し，I は加算するオクターブの範囲を表す．

$$c(k,t) = \sum_{f=0}^{I-1} H(12f + k, t) \tag{1}$$

また，通常の短時間 FFT による時間周波数解析では，低い周波数で十分な周波数分解能を得るためには広い窓幅が必要となるが，一方で，周波数分解能が比較的必要とならない高い周波数での時間分解能を下げる問題が生じる．そこで，周波数と窓幅の比を一定に保つことができる定 Q フィルタバンク解析では，高周波数での時間分解能を維持しながら，低周波数での分解能を確保することができる利点がある．本研究では，定 Q フィルタバンクを用いて時間周波数解析を行い，クロマベクトルを計算する．

図 1 に音響信号からクロマベクトル系列を作成する過程の概略を示す．このクロマの時系列ベクトルを和音認識における入力データと考える．

2.2 HMM による和音進行のモデル化

一般的に，調性音楽では和音の進行から楽曲が作成されると考えることができるため，和音の進行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．

ある和音が継続する区間のメロディは多様であるが，和声学に従ったものであるなら出現する音名には一定の傾向があるといえる．従って，音響信号から得られるクロマベクトルの系列にも同様のことが言え，和音に依存して出現しやすい音名の組み合わせは，クロマベクトルの分布によって表現することができる．

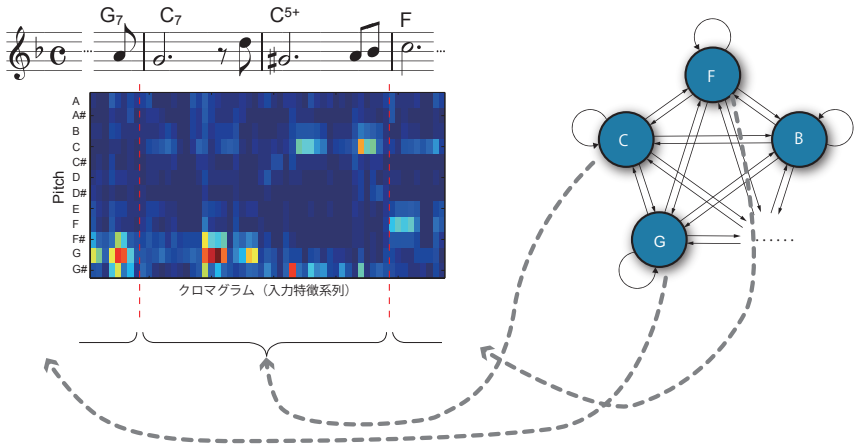

図 2 和音 ergodic HMM の概要
Fig. 2 An outline of ergodic chord HMM

また，和音の遷移の傾向についても和声学に基づいて定めることができる一定の規則があり，そのような規則を多くの事例（楽譜）から学習する確率モデルを考えることができる．これらの特徴を考慮したモデルとして，自動和音認識には HMM がよく用いられている．

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 によって和音進行をモデル化することを考える (図 2)．ここで，各状態では和音に依存したクロマベクトルの出力確率を持つこととする．なお，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，本来は調ごとに区別したモデルを作ることが望ましいが，ここでは簡略化して全ての調を 1 つのモデルで扱うことにする．

3. 環境依存 HMM による音響モデルの詳細化

HMM を用いた和音認識にはいくつかの先行研究があり，クロマベクトルの出力確率分布に着目して整理する．H. Papadopoulos らは，音響信号から得られるクロマベクトルから各状態における出力確率分布を直接学習せず，一定の規則に基づいてクロマベクトルの各次元の重みを定めている[4]．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が得られることが示されている．一方で，クロマベクトルから出力確率分布を学習するアプローチとしては，内山らはクロマベクトルはその時刻の和音のみに依存して確率的に生成されると仮定し，各和音のクロマベクトル列から出力確率分布を学習している[5]．また，上田らは典型的な和声パタンを語彙と考え，連続する和音に依存したモデル化を試みている[6]．

ところで，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ベクトルのパターンは複雑な分布を持つことが考えられる．先に示した先行研究では，クロマベクトルの傾向は特定の和音だけに依存したものとして扱われているが，本研究では，クロマベクトルの出現傾向をより精密にモデル化することの効果について検討する．

3.1 前後環境を考慮したモデル分類

ある和音から観測されるクロマベクトル系列は，その和音だけでなく，前後の和音進行にも影響を受けていると仮定し，音声認識の分野でよく用いられる前後の音素環境を考慮した triphone モデルの考え方を和音認識に導入する．このようにして，本研究では和音の分類を 24 種類として扱うが，その前後の連鎖に依存して場合分けをすることにより，理論的な分類は $24^3 = 13,824$ 通りのモデルを考えることになる．

なお，本論文では和音連鎖の表記法として，当該の和音を中心にして “–” を挟んで先行する和音，“+” に続いて後続する和音を表すこととする．

3.2 決定木に基づく HMM のパラメータ共有化

前後の和音を考慮してモデルを詳細に分類することにより，モデルの精度向上が期待されるが，そこにはいくつかの問題が生じる．まず，すべての和音連鎖の分類に対して網羅的に学習データを用意できないという，学習データ不足の問題がある．それに関連して，仮に学習データが存在したとしても十分な学習ができなければ，過学習によって汎用的な和音連鎖モデルを推定できない問題が考えられる．先に述べたように，前後の連鎖を考慮するだけで 1 万種類以上もの和音連鎖を扱うことになり，これだけの分類をカバーするための学習データは現実的に難しいといえる．

本研究では，限られた学習データを有効に利用しつつ，詳細な分類でも和音連鎖モデルを学習するための手法として，各モデル間でパラメータを共有することを考える．これまでに，モデルパラメータを共有化する試みとして，学習された和音 HMM の各状態パラメータ空間において，近傍のパラメータを共有化することを検討してきた[7]．しかし，この手法では，和音連鎖のように詳細に分類され，学習データに存在しない未知の和音連鎖を認識する場合には，適切なモデルを事前に用意しておくことができない問題が生じる．

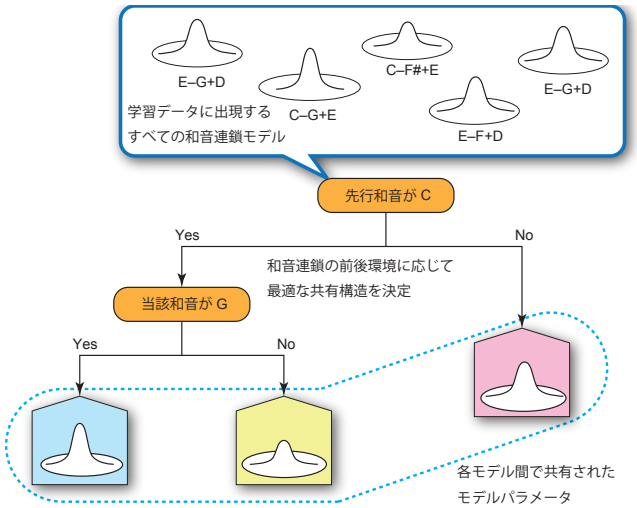

図 3 決定木に基づく HMM のパラメータ共有化
Fig. 3 An outline of parameter tying using tree-based clustering

そこで，既存の和音連鎖モデル間では，パラメータを共有化することで頑健なモデルを学習する一方で，未知の和音連鎖に対しても適切なモデルを選択可能な枠組みとして，図 3 に示す決定木に基づいたクラスタリング手法を導入する．

4. 実　　験

4.1 実験条件

提案手法の評価実験として，音楽 CD に含まれる音響信号から，和音連鎖 HMM を学習する．The Beatles のアルバム “With the Beatles” の CD から，各トラックの波形データをモノラル化し，ダウンサンプリングして用いた．表 1 に示す分析条件のもとで，定 Q フィルタバンクを用いて時間周波数解析を行った．さらに，フレーム t で得られた 12 次元のクロマベクトル $\boldsymbol{c}_t$ を静的特徴として，式 2 で計算される動的特徴 $\Delta\boldsymbol{c}_t$ を加えた合計 24 次元からなる特徴ベクトルを学習に用いる．

$$\Delta\boldsymbol{c}_t = \frac{\sum_{\theta=1}^{2} \theta(\boldsymbol{c}_{t+\theta} - \boldsymbol{c}_{t-\theta})}{2\sum_{\theta=1}^{2} \theta^2} \tag{2}$$

表 1 実験データと分析条件
Table 1 Details of experimental condition

楽曲	“With the Beatles”(The Beatles) の 14 曲
サンプリング周波数	11,025Hz (モノラル)
フレーム長	100msec
低 Q フィルタ分析のオクターブ数	6 オクターブ: 55.0Hz(A0)〜3,729.3Hz(Bb6)
特徴ベクトル	12 次元クロマベクトル +Δ クロマベクトル

和音の語彙は major，minor の 24 種類とし，それ以外の和音は第 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分布は，単一の多次元正規分布とし，クロマベクトルの各次元間の相関を考慮しない，対角共分散行列とした．

アルバム内の 14 曲中 13 曲を学習データとして用い，それぞれの学習セットの中で出現する和音連鎖について HMM を学習し，前後の和音に関する分類に基づいた決定木クラスタリングによって状態パラメータを共有化させる．その際，クラスタリングの停止基準を調整し，段階的に決定木の規模を変えながらモデルを作成した．本実験の HMM の学習には HTK(HMM ToolKit) を用いた．

4.2 パラメータ共有によるモデルの評価

クラスタリングによるパラメータ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それぞれのモデルを評価する．先の実験条件の下で，クラスタリングの条件を変えながら木構造のサイズを調整したモデルを用い，未学習データに関する Viterbi アライメントをとった際の尤度と，その際のモデルの規模の指標となるパラメータ数をプロットしたものを図 4 示す．

図 4 では，クラスタリングにおける分割の停止条件に応じて，木構造の作成が抑制されるため，モデルの規模を示すパラメータ数が単調に減少していることが確認できる（図中の青線）．モデルの規模の下限は，木構造をまったく分割しない状態ことを表し，その場合は前後環境によるモデル分類をまったく考慮しない単独の和音モデルに相当する．また，モデルの規模の上限は，学習データに出現したすべての和音連鎖を完全に分類した場合に相当する．

尤度のグラフ（赤線）を見ると，モデルの規模が大きい場合と小さい場合にそれぞれ尤度が低くなり，モデルの規模が 30〜40 になる中央付近にピークが存在することが分かる．モデルの規模が大きい場合には，各和音連鎖について詳細なモデル化がなされるが，未学習データに対する汎用性が低いため，尤度が下がる傾向があり，モデルの規模を小さくした場合では，いくらか改善する傾向は見られるものの，和音連鎖のモデル分類をまったくしない

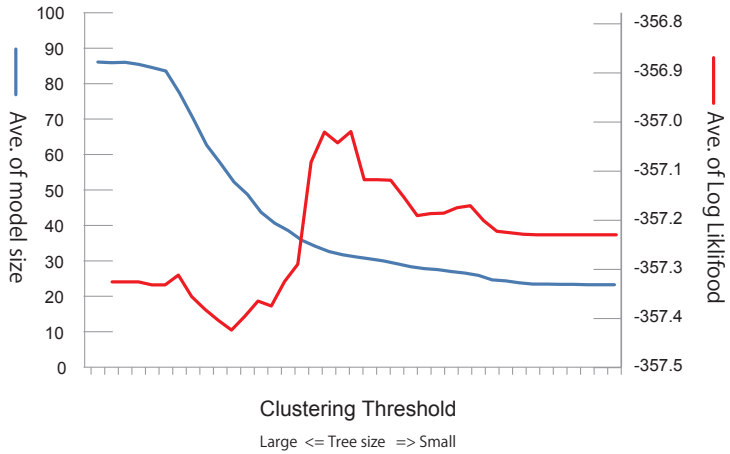

図 4 各モデルの Vitrebi パスの尤度
Fig. 4 Viterbi score of each model

ことで，モデルが平滑化され精度が低下していることが考えられる．一方でモデルの規模が 30〜40 付近では，適度にパラメータの共有化がなされることで，未学習データに対しても適合しやすいモデル化がなされていることが示されている．

4.3 作成された木構造

図 5 にクラスタリングによって得られた木構造の一例を示す．各和音モデルによって木構造の大きさが異なるのは，学習データに出現する各和音の出現数の違いによるものや，和音連鎖の影響を受ける度合いによる違いなどが考えられるが，何らかの尺度によって最適な木構造の大きさを決定する必要がある．具体的には，MDL(Minimum Description Length) などの情報量基準を用いてクラスタリングの停止基準を定める手法が利用できる．

5. む す び

本稿では，HMM に基づいた音楽音響信号からの自動和音認識において，音響特徴をより詳細にモデル化する試みとして，和音の連鎖ごとに音響モデルを詳細に分類し，そのようなモデルを効率的に学習するためのクラスタリング手法を検討した．これは，音響信号から観測されるクロマベクトルが，その時刻の和音だけでなく，周辺の和音からなる和音進行からも一定の影響を受けていると考え，和音の連鎖モデルとして前後の和音連鎖による

Vol.2010-MUS-88 No.4
2010/12/4

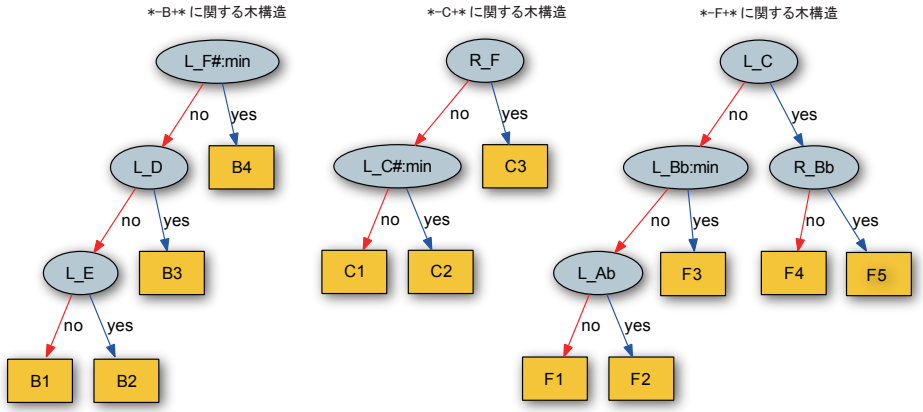

図 5 作成された木構造の例 (中心和音が B,C,F の場合)
Fig. 5 An examples of generated tree structure

trigram モデルへ拡張することで性能の向上を図った．

モデルの分類を詳細化することにより学習データ不足などの問題があり，頑健なモデルを学習するための方法として，類似したモデルパラメータをもつ状態を共有することを検討した．また，モデルの分類が多数に上ることから，未学習の和音連鎖に対処するためのクラスタリング手法として，木構造によるクラスタリング手法を導入した．

評価実験では，モデルの木構造の規模を変化することで和音連鎖を考慮しない場合に比較して，提案手法の有効性を確認した．また，木構造を適度に調整し，モデルパラメータの共有度合いの最適にすることで，音響モデルの詳細化によって和音認識性能を向上させるという期待が示された．

今回は音響モデルによる評価を行ったが，和音遷移モデルを併用して，音響信号からの和音認識をすることが課題である．また，今後の展望として，音響信号からの自動和声認識と，これまでに提案した楽曲間類似度の評価手法[1] を組み合わせることで，音響信号から類似曲検索を行う手法を展開する．

謝辞 本研究の一部は，財団法人電気通信普及財団 平成 21 年度研究調査助成による支援を受けて行われたものである．

参 考 文 献

1) 伊藤 綾, 酒向 慎司, 北村 正, “コード進行クラスタリングによる楽曲のモデル化と楽曲間類似度の評価”, 第 8 回情報科学技術フォーラム, E-037, pp.341–342, 2009.
2) Alexander Sheh and Daniel P.W. Ellis, “Chord segmentation and recognition using EM-trained hidden Markov models”, *Proc. of International Conference on Music Information Retrieval (ISMIR)*, pp.183–189, 2003.
3) Takuya Fujishima, “Real-time chord recognition of musical sound: A system using common lisp music”, *Proc. International Computer Music Cofference (ICMC)*, pp. 464–467, 1999.
4) Hélène Papadopoulos and Geoffroy Peeters, “Large-scale study of chord estimation algorithms based on chroma reprentation and HMM”, *Proc. of Content-Based Multimedia Indexing (CBMI) ’07*, pp.53–60, 2007.
5) 内山 裕貴, 宮本 賢一, 西本 卓也, 小野 順貴, 嵯峨山 茂樹, “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.
6) 上田 雄, 小野 順貴, 嵯峨山 茂樹, “機能和声モデルによる音楽信号の和声推定”, 情報処理学会研究報告–音楽情報科学 (MUS), vol.2010-MUS-86, no.13, pp.1–6, 2010.
7) 伊藤 綾, 酒向 慎司, 北村 正, “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”, 第 9 回情報科学技術フォーラム, 第 9 回情報科学技術フォーラム, E-029, pp.285–286, 2010.