

隠れマルコフモデルを用いた音響信号からのコード認識と楽曲分類に関する研究

代表研究者 酒 向 慎 司 名古屋工業大学大学院 情報工学専攻 助教

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規模化によって、多量の楽曲を手軽に扱えるようになった。これに伴い、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、強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の様々な研究が展開されている。例として、リズムを入力することで楽曲検索を行うシステム[1]や、楽曲の音高や音長、旋律パターンを用いた類似曲検索[2]などが研究されている。

ここで挙げた研究には、リズムやメロディ等が用いられていたが、これらの音楽的な特徴量には、人が手動でラベリングしたものやMIDI形式のデータが用いられている場合が多い。リズムやメロディ等の音楽的要素を音響信号から精度よく抽出する試み[3,4]は多数なされているが、多重音を扱うものになると、未だ難しい問題である。これらの試みは主に自動採譜を目的にしたものであるが、楽譜のレベルまで詳細な情報を得ることができなくとも、楽曲の雰囲気や掴むことは可能であると考えた。例えば、音楽の専門的な知識がない人でも、音楽を楽しみながら楽曲の違いや類似性を見出すことができる。これは音楽が詳細な音符の並びとして知覚されているというより、音符の集合がどのようなものであったか、というところに注目していると考えることができる。従って、楽曲を簡略化することによって、楽曲の特徴を数量化できれば、楽曲の類似性を直接的に比較することができ、類似曲検索などの技術へ応用することが可能となる。

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[5]。ここでは、楽曲というものが楽的に意味のあるフレーズによって構造的に作られていることから、和音進行を部分的に比較し、その累積を比較することで楽曲間の類似度を得られると考えた。この手法でははじめに、楽曲を意味のあるまとまりに分割し、それぞれの和音進行間の類似度を、近親調や、和音の構造から算出する。次にこの類似度をもって和音進行をクラスタリングすることで楽曲をモデル化する。最後にモデルのクラスタ出現頻度に関するヒストグラムを比較することで、楽曲間の類似度を得たしかし、この手法では楽曲の和音進行が既知である場合のみを扱ったため、音楽音響信号から直接類似度を求めることはできなかった。そこで、音響信号からの自動和音認識を組み合わせることで、音楽音響信号の類似性を直接比較することを目指す。本研究では、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。

2.和音進行による楽曲比較

楽曲は音楽的に意味のある小さなまとまりに分解することができ、その繋がりにも音楽的な意味付けがなされていると考えられる。様々な楽曲に共通するこの意味のある要素を得られれば、その構成や順序関係によって楽曲を比較することができると考えた。今回は、要素の累積から類似度を評価する方法を考える。このような要素の共通性を見出す特徴として、楽曲の大局的な流れを把握するために有効と考えられるコード進行を用いると、楽曲間で共通する意味のまとまりというものは、様々な楽曲に現れるコード進行パターンであると考えられる。したがって、楽曲のコード進行をカデンツのような一連のパターンに分解し、それを

分類することにより楽曲間で共通している部分を見つけ出すことができる。

ここでは、一般的に音楽で1つのまとまりとして区切りが良いとされる4小節ごとに楽曲を分解し、これをブロックと呼ぶことにする。様々な楽曲から得られたブロックの集合をクラスタリングすることで代表的なコード進行パターンを求め、各ブロックをこれに置き換えたものを用いて類似度を比較する。

2.1 コード進行の分類

コード進行のデータ集合から共通要素を見出す方法として、長澤らのコード進行類似度[6]を用いる。和音の進行には原則や禁則が存在する。これにより、和音進行には代表的なパターンが存在し、近傍にある和音進行はそのパターンをアレンジしたものと考えられる。例えば、ハ長調の代表的な和音進行である「C→G→C」は、Cの代わりに平行調の関係にあるAmを用いて「C→G→Am」とアレンジされることがある。このような和音進行を類似したものとして扱うため、和音進行をクラスタリングすることで楽曲のモデル化を行う。先行研究ではk-means法が用いられていたが、ランダムに設定される初期値に結果が大きく左右されるため、ここではLBGアルゴリズムを採用する。

LBGアルゴリズムとは、k-means法から派生した分割最適化クラスタリング手法であり、歪み最小基準でクラスタリングを行う。クラスタ中のコード進行パターンのばらつきが大きい場合は、全く異なるコード進行が類似していると見なされ、ばらつきが極端に小さい場合には互いに類似しているコード進行パターンが別々のクラスタを形成している可能性がある。このようなクラスタが生成されることを避けるため、すべてのクラスタ内の分散が一樣になることが好ましいと考え、先行研究ではクラスタ分割の停止条件としてクラスタ内分散の閾値を設定した。しかし、クラスタ内分散が閾値以下となるクラスタでも、分割した結果、新たな2つの中心間の距離が大きくなる場合には分割を行った方が好ましいと考えられる。そこで本研究では、分割の基準としてクラス内クラス間分散比を考える。以下にそれぞれを説明する。

従来法:クラスタ内分散によるクラスタリング

クラスタ数を k 、 i 番目のクラスタを C_i 、クラスタ C_i 中のブロック数を n_i 、ブロック a とブロック b の距離を $D(a,b)$ 、クラスタ C_i の中心を $medoid(C_i)$ 、クラスタ C_i 中の j 番目のブロックを $C_i(j)$ とし、クラスタ C_i のクラスタ内分散 $V(C_i)$ を次式で定義する。クラスタ内分散が閾値以下となったクラスタはそれ以上分割しない。この分割方法により、比較的分散の偏りが小さいコードブックが得られる。

$$V(C_i) = \frac{1}{n_i} \sum_{j=0}^{n_i-1} D(\text{medoid}(C_i), C_i(j))$$

提案法:クラス内クラス間分散比によるクラスタリング

全体のブロック数を n 、全ブロックの中心を $medoid$ で表し、クラス内分散 σ_w 、クラス間分散 σ_B 、クラス内クラス間分散比 J_σ を以下の式で定義する。

$$\sigma_w = \frac{1}{k} \sum_{i=0}^{k-1} V(C_i)$$

$$\sigma_B = \frac{1}{n} \sum_{i=0}^{k-1} D(\text{medoid}, \text{medoid}(C_i)) \times n_i$$

$$J_\sigma = \frac{\sigma_B}{\sigma_w}$$

クラス内クラス間分散比が最大となるクラスタを順次分割していく。したがって、クラスタ内分散をより小さく、それぞれのクラスタ間の距離をより大きくするコードブックが得られる。クラスタリングを行うためのコード進行間の距離尺度には編集距離を用いる。コードは例えば C_m のように、根音を表す $C, C^\#, D^b, \dots, B$ と他の構成音の音程を表す M, m, \dots, m から成っている。それぞれについて、近親調、コード間の関係から編集コストを定義し、編集距離でコード進行間の類似度を求める。このようにして、共通の代表コード進行に置き換えられた楽曲を比較することで楽曲間の類似度を得る。ここでは要素の順序関係は考慮しないため、クラスタの出現頻度に関するヒストグラムのユークリッド距離を楽曲間の距離とする。

3. 音響信号からの和音認識

和音認識の研究では、和音進行の音楽的制約と音響的特徴を隠れマルコフモデル(HMM)の枠組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。

3.1 クロマベクトルによる音響特徴抽出

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

(Chroma 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性が示されている。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。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。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。

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

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)。

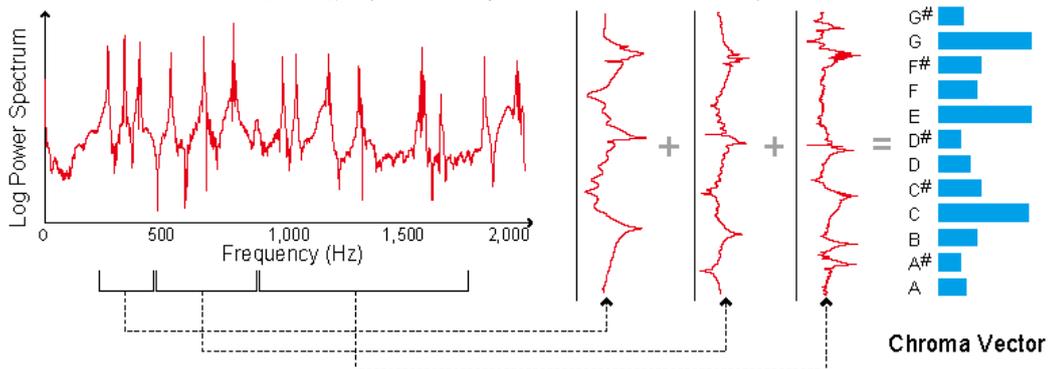


図1. クロマベクトルの概要

3.2. 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。

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。

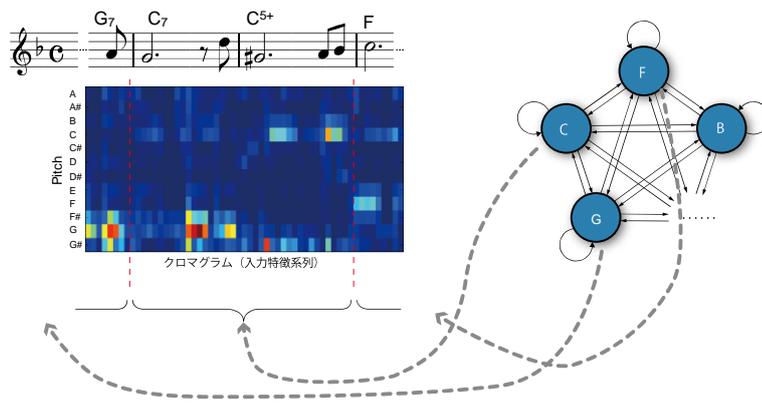


図 2. 和音 ergodic HMM の概要

3.3. 和音モデル間の共通性

和音から作られるメロディは多様であること使用楽器により音響特徴が変化すること、非定常音が含まれること等の影響により、実際に1つの和音から生成されるクロマは複雑で多様である。従来は、音楽的知見に基づいた和音間の共通性を利用することで単独和音の音響特徴を厳密に学習しない手法が有効であった。しかし、この共通性が本当にそのデータに適しているか調べることはできないため、本研究では学習データから自動で和音間の最適な共有構造を得る手法を提案する。

本稿では、同一和音区間での音響特徴の変化がその前後の和音進行から影響を受けていると仮定し、和音モデルを前後の連鎖に依存して場合わけをすることで音響モデルを細分化する。

そして出現する和音連鎖モデルを、前後の和音の一致をみて木構造でクラスタリングすることで統合し、最適な共有構造を得る。

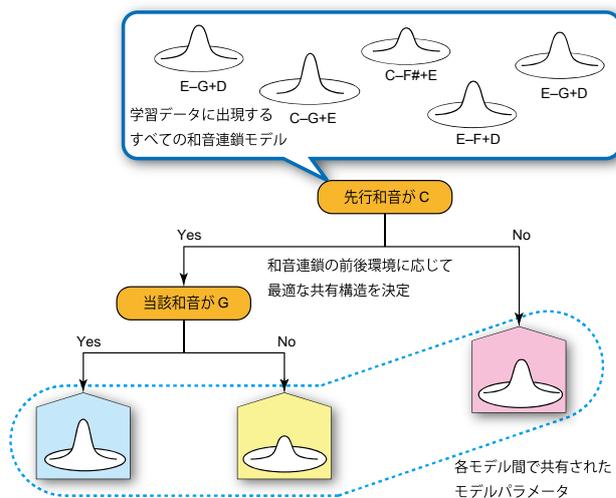


図 3. 決定木に基づく HMM のパラメータ共有化

4. 評価実験

4.1 和音進行による楽曲分類

2つのクラスタリング手法の性能を評価するため、聴取実験によって得られた主観的な類似度との比較を行う。

実験条件

楽曲間の主観的な類似度を得るため聴取実験を行う。基準曲1曲に対し比較対象曲5曲を用意したものをデータセットとし、どの曲が基準曲に類似しているか順位付けをする。実験では基準曲と比較対象曲中の2曲を1組として提示し、基準曲に近いと感じた方を選択させた。データセットを3つ用意し、20人の評価者にそれぞれ15組分を評価させ、1組につき10個のデータから類似度を得た。

楽曲のコード進行データは ultimateGuitar.com[10]から、4拍子のポピュラー音楽を対象に収集した、

Dance, HipHop, Pop, Punk, R&B, Rock 等のジャンルを含む 50 曲を用いた。また、ここでは 2 つの分割停止条件を比較するためクラスタ数を 8 と定め、そのようにクラスタ内分散の閾値を設定した。聴取実験では 3 ブロックのみを用いるため、1 曲の中からサビ付近 3 ブロック分を切り出してモデル化した。

実験結果

クラスタリングの様子を図 1 に示す。図 4(a) はクラスタ内分散の閾値を 155 に設定した場合、図 4(b) はクラス内クラス間分散比を最大とするクラスタから順次分割した場合を示しており、ノード下の数字は分割の順序を表している。また、これらのコードブックを用いてモデル化した楽曲の客観的距離と、聴取実験によって得られた主観的類似度との関係を図 5 に示す。

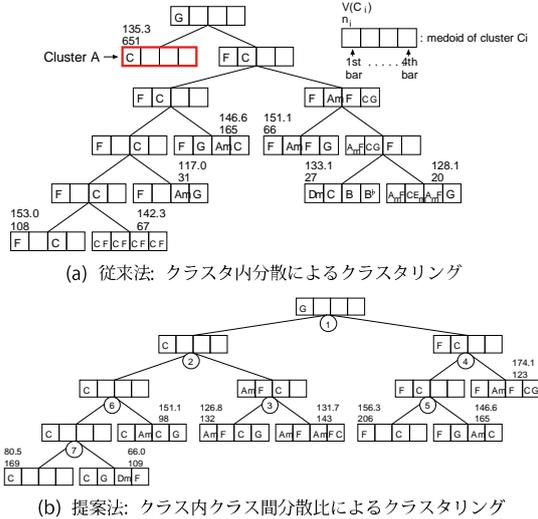


図 4. クラスタ分割の様子

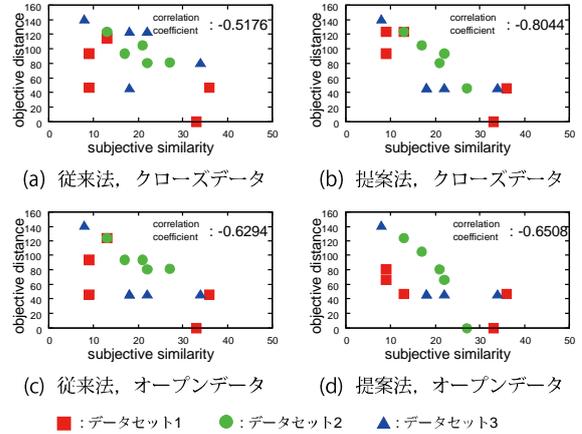


図 5. 客観的距離と主観的類似度の関係

考察

図 4 より、2 つの手法でクラスタリングの過程が大きく異なることがわかる。例えば、図 4(a) では末端となるクラスタ A は図 1(b) では 5 クラスタに分割されている。クラスタ A の medoid は頻出するコード進行であるため、他の類似していないブロックが含まれていたとしてもクラスタ内分散は小さくなる傾向にあることが原因である。この結果から、コード進行をクラスタリングする際のクラスタ内分散は、そのクラスタの最適性を示しているとは限らないと言える。

また、図 5 から、(a)、(c) と比較して (b)、(d) の方が、客観的距離と主観的類似度との相関が高くなることがわかり、クラスタ内分散を基準とする場合よりもクラス内クラス間分散比を基準とする場合で主観的類似度に近い類似度を得られることがわかる。

4.2 和音認識実験

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" から、各トラックの波形データをモノラル化し、11,025Hz にダウンサンプリングして用いた。フレーム長 100ms, A0 から 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。

和音の語彙は第 3 音に着目して major, minor に近似し、24 種類とした。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。

実験結果と考察

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、未学習データの認識率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、未学習データの認識率と、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 6 示す。図 6 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。

認識率のグラフにおいて、モデルの規模が大きくなった場合に認識率が下がっているのは、各和音連鎖について詳細なモデル化がなされているが未学習データに対する汎用性は低いと考えられる。また、モデルの規模を小さくした場合は、和音連鎖のモデル分類をしない単独和音モデルに近くなり、モデルが平滑化されているために精度は低下している。認識率の上がっている中間部分では、適度なパラメータ共有がされることで、未学習データに対しても適合しやすいモデル化がされたことが示されている。

また、同条件の単独和音モデルで認識を行った場合の認識率の平均は 31.14% であり、音響モデルを詳細化した効果が見られる。

ここで、音響信号からの和音認識結果には少なからず誤りが含まれるため、類似度算出に影響が出ることが考えられる。和音認識においては、共通音を多くもつ近親調の関係にある和音間での認識誤りが多い。しかし先に述べた手法では近親調などを考慮してクラスタリングを行うため、このように近い誤りは正解の和音進行と同じものとして扱われる。従って、ある程度の和音認識誤りは許容できると考えられる。

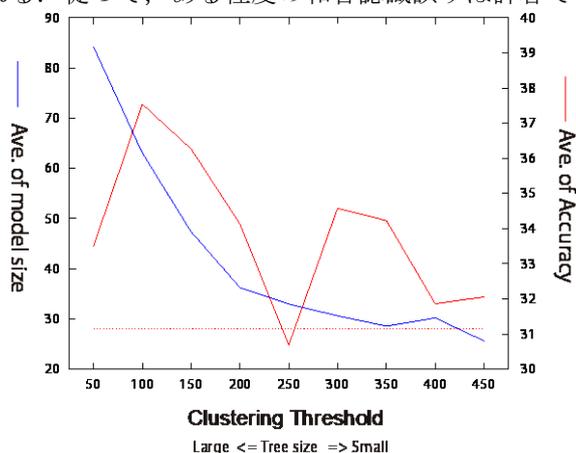


図 6. 各モデルの和音認識率

5. むすび

本研究では、音響信号を対象とした類似曲分類を実現させるための手法を提案し、その要素技術について検証を行った。一方は和音進行既知の場合に楽曲間の類似度を算出する手法である。和音進行の距離に基づいて楽曲のブロックをクラスタリングしたものを楽曲のモデルとし、それぞれの出現クラスタに関するヒストグラム間の距離を楽曲間の距離とした。聴取実験による主観的な楽曲類似度と比較したところ、同様の類似度が計算されることが確認された。もう一方は上で述べた手法を音響信号に対応させるため、音響信号から和音進行を認識する手法である。我々は従来から提案されている HMM に基づく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、観測されるクロマベクトルがその時刻の和音だけではなく、その前後の和音からも影響を受けていると考えたためである。そこで、単独の和音モデルを前後の和音によって分類する trigram モデルへ拡張した。またここで、モデルを詳細化することによる学習データ不足などの問題を解決するため、木構造によるクラスタリングを導入することで類似したモデルパラメータを共有することを検討した。評価実験では、木構造の規模を適度に調整し、モデルパラメータの共有度合いを最適にすることで和音認識率が向上した。

今回はこれら 2 つの手法を個別に検証したが、今後は 2 つを実際に組合せることで音響信号を分類し、聴取実験の結果と比較することでその有効性を確かめる必要がある。また、和音の認識率の低下による類似度への影響の調査や、和音進行のクラスタリングで用いた近親調の考え方を和音連鎖のクラスタリングに導入することなどが課題として挙げられる。

【参考文献】

- [1] 池谷 直紀, 服部 正典, 梅木 秀雄, 大須賀 昭彦: “リズム入力インタフェース「タタタタップ」による大規模音楽検索”, *IPSJ SIG Technical Notes*, Vol.2005, No.52, pp.27–33, 2005.
- [2] 辻 康博, 星 守, 大森 匡: “曲の局所パターン特徴量を用いた類似曲検索・感性語による検索”, *IEICE Technical Report. Speech*, Vol.96, No.565, pp.17–24, 1997.
- [3] 清水純, 丸山剛志, 三浦雅展, 柳田益造: “ハミングによる単旋律の自動採譜”, *音楽音響研資*, MA2004-49, 2004.
- [4] 菊地 淑晃, 後藤 真孝, 村岡 洋一: “ベースギターの自動採譜システム”, *情報処理学会全国大会講演論文集*, Vol.第 52 回平成 8 年前期, No.1(19960306), pp.459–460, 1996.
- [5] 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.
- [6] 長澤 慎子, 渡辺 知恵美, 伊藤 貴之, 増永 良文: “ポピュラー音楽クラスタリングのための近親調を用いたコード進行類似度の提案”, *IPSJ SIG Technical Report*, Vol.2007, No.37, pp.69–76, 2006.
- [7] 北川 祐: “ポピュラー音楽理論”, *リットーミュージック*, 2006.
- [8] Alexander Shehand Daniel P. W. Ellis, “Chordsegmentationandrecognitionusing EM-trained hidden Markov models”, *Proc. of International Conference on Music Information Retrieval (ISMIR)*, pp.183–189, 2003.
- [9] Takuya Fujishima, “Real-time chord recognition of musical sound: A system using common lisp music”, *Proc. International Computer Music Conference (ICMC)*, pp.464–467, 1999.
- [10] <http://www.ultimate-guitar.com/>

〈発表資料〉

題 名	掲載誌・学会名等	発表年月
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	2010 年 9 月
混合音の調波構造パターンとの相関による楽器・音高推定	2010 年電子情報通信学会基礎・境界ソサイエティ大会	2010 年 9 月
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	報処理学会音楽情報処理研究会	2010 年 11 月
音楽的知見を考慮した環境依存 HMM に基づく音響信号の自動和音認識の検討	第 10 回情報科学技術フォーラム	2011 年 9 月 (発表予定)