

関係調を考慮したHMMに基づく音響信号の自動和音認識と類似曲分類

杉山 雄一[†], 酒向 慎司[†], 北村 正[†][†] 名古屋工業大学大学院工学研究科

1 はじめに

音楽コンテンツの多様化・大規模化に伴い、類似曲検索システムのニーズが高まっている。従来の類似曲検索として、ユーザの行動履歴に基づく手法が広く利用されているが、ある程度の利用者が確保されなければ実用性に欠けるといった難点も存在する。

このような意識から、音響的な情報を用いて楽曲間の類似性を評価する手法 [1] が提案されてきている。用いられる音響情報として、メロディやリズムなどの詳細な特徴が考えられるが、音響信号からの抽出は未だ困難である点に加えて、楽曲間の比較方法に関しても検討の余地が残る。

そこで、楽曲間比較を簡素化する枠組みとして、楽曲のフレーズ構造に着目する。そもそも、楽曲は音楽的な意味を持つ様々なフレーズの集合として構成される。楽曲の構造をフレーズにより近似的に表現できれば、楽曲間比較の簡素化が可能になると予想される。楽曲構造を近似するため、音響情報に基づいてフレーズの種類を行いたい。ここで、各フレーズの音楽的意味は、詳細な音響特徴というよりはそれらを包含した大局的な特徴により表現されると考える。メロディ・リズム等が異なっても、フレーズの類似性を判断できることがよい例である。このような考えから、大局的な音響特徴を用いたフレーズの種類を考える。この大局的な音響特徴として、和音進行に着目する。和音進行は楽曲の雰囲気表現する一要素であることに加え、和声学に基づく一定の規則性があるため、それらを利用することで、柔軟に和音を推定することが可能となると考える。

以上の考えに基づき、本稿では柔軟な類似曲検索に向けて、音響信号から和音進行を推定しそれを基に楽曲構造を近似し客観的に楽曲間類似度を評価する枠組みを提案する。

2 システムの概略

システムの全体像を図1に示す。1. 音響信号から自動和音認識を行い、2. 認識和音をフレーズに分割する。次に、3. 和音進行によりフレーズを自動分類し、4. 楽曲ごとにフレーズのリストグラムをとる。最後に、5. 楽曲間でヒストグラムを比較することにより客観的な距離を算出する。

また、和音に関する知見として、関係調の考え方が存在する。和音間の相対的な親密さを表現したもので、五度圏 [2] における各和音間の配置の近さで定義される。この関係性を数量化したものとして、編集距離 [3] という距離尺度が提案されている。この考えを利用し、和音の認識・フレーズの種類性の自動決定を行う。

3 音響信号の自動和音認識

3.1 和音に関する特徴量: 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じであれば同一の和音として認識される。転回形を考慮した特徴量として、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クロマベクトル [4] が有効であると考えられる。これを和音に関する特徴量として、音響信号から抽出する。

3.2 HMM による和音進行のモデル化

調性音楽は和音進行に基づいて作曲されると仮定する。和声学に基づくと、和音を構成する音名の出現には一定の傾向があるといえるため、クロマ系列で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できる。また、和音間の遷移の傾向においても和声学に基づく一定の規則性があり、確率的に扱うことができる。これらの特徴を考慮した和音進行のモデルとして、隠れマルコフモデル (HMM) がよく用いられる [4]。自動和音認識は HMM を用いて、観測されたクロマ系列から和音進行を求める逆問題として定式化できる。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 で和音進行をモデル化する。また簡略化のため、全ての調を 1 つのモデルとして扱う。

3.3 和音モデルの共通性による状態共有

和音を構成する音名は一律ではないことや様々な非定常音が含有することから、各和音から生成されるクロマは多様であり複雑となる。本稿では音響特徴を精密にモデル化する枠組みとして和音区間における音響特徴が前後和音に依存すると仮定し、和音のモデルを前後和音に対して細分化することで音響モデルの精密化を図る。この際に生じる、モデルの学習データ不足やモデルの汎用性の低下などといった問題に対し、決定木に基づくクラスタリングを行うことで、共通性をもつモデルパラメータを共有し、頑健さを維持する。その際に、関係調に基づく分類基準を定めることで、和音を扱う上で重要な和音間の相対的な関係性を考慮した [5]。

4 楽曲構造の近似と楽曲間比較

楽曲のジャンルや楽曲構造ごとに、出現するフレーズに傾向が見られることから、楽曲中には一定の共通性を持った類似したフレーズのパターンが存在すると思われる。この共通性を、大局的な演奏情報である和音進行によって決定できると仮定し、各フレーズの認識結果を用いて自動分類を行う。フレーズの空間における絶対的な配置は容易ではないが、関係調の知見を用いることで、フレーズ間の相対的な位置関係は計算可能ではと考えた。そもそも、類似性の高いと認知されるフレーズ間においては、和音進行に関しての局所的なアレンジがなされていると考えられるので、それらを吸収可能な距離尺度として、編集距離をコストに用いた DP マッチングにおける累積距離を用いる。フレーズ A, B の認識和音系列を $A = \{a_1, a_2, \dots, a_N\}$, $B = \{b_1, b_2, \dots, b_M\}$, $d(i, j)$ を a_i と b_j の編集距離, $g(1, 1) = d(1, 1)$ としたとき、フレーズ間の距離 $g(N, M)/(N + M - 1)$ を式 (1) で算出する。

$$g(i, j) = \min \left\{ \begin{array}{l} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i, j-1) + d(i, j) \end{array} \right\} \quad (1)$$

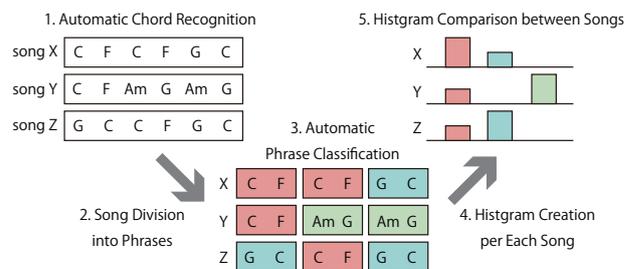


図1: システムの全体像

表 1: 実験データと分析条件

| | |
|-----------|----------------------------|
| 楽曲 | The Beatles の全 189 楽曲 |
| サンプリング周波数 | 11,025 Hz(モノラル) |
| フレーム長 | 100 msec |
| オクターブ数 | 55.0 Hz(A0)~3,520.0 Hz(A6) |
| 特徴ベクトル | 12 次元クロマベクトル+ Δ |

各フレーズ間の相対距離を用い、クラス内クラス間分散比最大化基準において、LBG アルゴリズムによりフレーズの自動分類を行う。

楽曲中での各フレーズの出現傾向が楽曲らしさを表現するという考えより、楽曲ごとに出現フレーズの統計を取り、それらのユークリッド距離により客観的距離を算出する。

5 評価実験

5.1 和音認識実験

和音特徴の生成モデルとして、和音連鎖 HMM を学習する。楽曲は CD のデータをモノラル化し、ダウンサンプリングしたものを用いた。表 1 に示す実験条件のもと、定 Q フィルタバンクによる時間周波数解析を行い、音響信号からクロマベクトルを抽出した。HMM における各和音の出力確率は単一の多次元正規分布と定め、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。和音の語彙は major, minor の計 24 種類とし、それ以外の和音は第 3 音に着目して major, minor へ近似した。

学習データ中に出現する各和音連鎖において HMM を学習したのちモデルのパラメータ共有を行い、モデルにおける Viterbi 探索により、認識データに対して事後確率が最大となる和音系列を推定する。この際、クラスタリングの停止基準を調節し、段階的に決定木の規模を変化させながらモデルを作成した。全楽曲を学習・認識に用いる場合をクローズセット、発表順における偶数番アルバムを学習に、奇数番アルバムを認識に用いる場合をオープンセットとし、各々において和音の認識率を算出する。

5.2 主観評価実験

得られた楽曲間の客観的距離の妥当性を評価するため、聴取実験により主観的類似度を測定し、それらの比較を行う。

本研究では、楽曲構造ごとに楽曲を分割した各々を類似度の比較対象とする。各比較対象の曲中において、4 拍からなる逐次 4 小節を 1 フレーズと定め、和音の認識結果を用いて、クラスタ数 16 を停止基準としフレーズのクラスタ分割を行う。なお、楽曲構造情報、拍情報 [6] は既知とし、和音情報はクローズセットにおける最良な認識結果を用いた。

基準曲 1 曲と、対する客観類似度が段階的な 5 曲の比較曲、計 6 曲を 1 つのデータセットとし、どの比較曲が基準曲に類似しているかを順位付けする。実験では、基準曲と比較曲中の 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 11 人の各評価者に全 30 組分を評価させ、1 組につき 11 個のデータより類似度を得た。

5.3 結果と考察

和音認識実験に関して、図 2(a) にオープンセットにおける認識結果を、図 2(b) にクローズセットにおける認識結果を示す。いずれも音響モデルの精密化により、認識率の向上が確認できる。オープンセットに関して、モデルサイズ 145 の際に 32.87% の認識率が得られた。モデルサイズにより認識率に揺らぎが生じているが、モデルサイズが大きくなるに従い、認識率が低下する傾向にある。これは、過度なモデルの詳細化により汎化性が低下したためだと考えられる。クローズセットにおいては、モデルサイズ 1056 の際に 60.08% の認識率が得られた。オープンセットとは対称的にモデルを細分化するに従い、認識率が向上する傾向が見られた。これは、認識データの特徴を学習データで表現できているため、モデルの詳細化が十分に機能した結果であるといえる。

また、主観評価実験に関して、得られた客観的距離との関係性を図 3 に示す。データセットによりばらつきはあるが、総合的に一定の相関が見られたことから、客観的距離が人間の感性と繋がりが持つことが確認できる。データセット 1 に関しては相関があまり見られなかったが、聴取時に抱く印象として、ジャンルやリズムなど和音進行以外の影響も受けているからであると予想される。

6 まとめ

本稿では、楽曲のフレーズ構造とその和音進行に着目し、客観的に楽曲間の類似度を評価する枠組みを提案した。音響信号から推定した和音進行を用いてフレーズの共通性を自動決定し、楽曲構造を近似したのちに、楽曲間の出現フレーズの傾向の違いにより客観的距離を算出した。聴取実験の結果、客観的距離と主観的類似度に一定の相関がみられたことから、本手法の有用性が示唆された。

今後の課題として、和音認識に関しては調情報を考慮してモデルを作成すること、類似曲分類に関しては、3 拍子の曲や不規則性などフレーズとして扱う幅を広げること、リズムやジャンルなどの和音進行以外の特徴も考慮して客観評価を行うことなどが挙げられる。また、双方の連結に関して、和音の認識率低下による客観的類似度への影響の調査も課題として残る。

参考文献

- [1] 大野ら, “楽曲全体における特徴量の傾向に基づいた類似検索手法”, 日本データベース学会論文誌, 7(1), pp.233-238, 2008.
- [2] 北川 祐, “ポピュラー音楽理論”, リットーミュージック, 2006.
- [3] 長澤ら, “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”, 情報処理学会研究報告, 2007(37), pp.69-76, 2007.
- [4] Alexander Sheh *et al.*, “Chord Segmentation and Recognition using EM-trained Hidden Markov Models”, *Proc. of ISMIR*, pp.183-189, 2003.
- [5] 杉山ら, “関係調を考慮したパラメータ共有 HMM に基づく音響信号の自動和音認識の検討”, 情報科学技術フォーラム講演論文集, 10(2), pp.303-304, 2011.
- [6] <http://www.isophonics.net/>

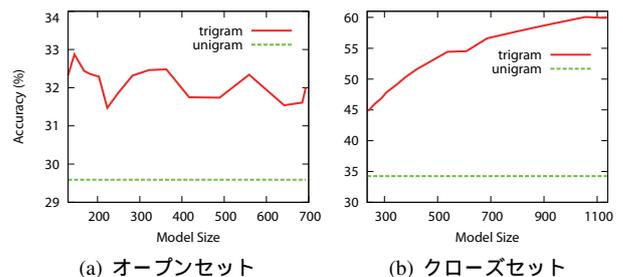


図 2: 各モデル規模における和音認識結果

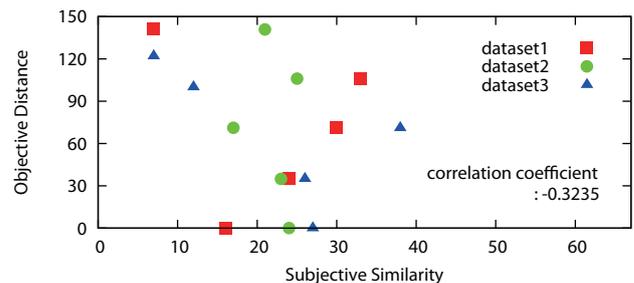


図 3: 客観的距離と主観的類似度の関係性