1P-06

隠れマルコフモデルを用いた楽曲の旋律と歌詞の自動対応推定

鏡味　なつみ†　　岩野　公司†
東京都市大学†

1. はじめに

我々は，大量の楽曲データを対象とした音楽分析の一手段として，「楽曲の旋律（メロディ）と歌詞を読み上げた際の声の高低のパターンの類似度を自動分析するシステム」の提案と構築を行っている[1]．このシステムでは，まず，独立に入力された旋律（音符・休符情報）と歌詞（テキスト）に対し，歌詞のどの部分がどの音符に対応付けられるかを自動推定する．先行研究[1]では，この対応付けに DP マッチングを利用した手法を提案し，利用している．この方法では，DP マッチングにおけるスコアを「歌詞のモーラの種類」「休符の長さ」といった情報に基づくルールによって制御している．しかし，従来手法では，特定の楽曲で十分な推定性能が得られず，その理由として，音符の長さや高さの情報を利用していないことが挙げられる．

そこで本研究では，音符の長さや高さの情報の取り込んだ自動対応推定手法として，隠れマルコフモデル（Hidden Markov Model: HMM）を利用した手法の提案を行う．具体的には，各モーラを HMM でモデル化した上で，音符の詳細な情報を反映させた特徴量を入力して最尤となる状態遷移系列を求めることで，両者の対応を推定する．この手法により情報量を増やすことによる対応推定性能の改善を図る．

2. 従来手法における旋律と歌詞の対応推定手法

旋律と歌詞の対応推定では，まず，歌詞のテキストを「ももたろうさん、」のように，読点や改行記号を含むモーラ・記号系列に変換した上で，それぞれに対して，旋律中のどの音符・休符に対応するのかを決定する．

従来の DP マッチングを用いた対応推定[1]では，まず，入力のモーラ・記号系列と音符・休符系列をそれぞれ横軸と縦軸に配した 2 次元空間を構成する．各対応点からの上，右，斜め右上方向モーラ－1 音」の 3 種類の対応を表すことになる

Automatic alignment of music melody and lyrics using Hidden Markov Models
Natsumi Kagami†, Koji Iwano†, †Tokyo City University

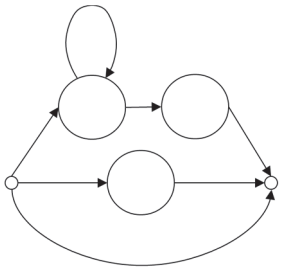

図 1　対応推定に用いるバイモーラ HMM の構造

の遷移が「1 モーラ－2 音」「2 モーラ－1 音」「1 ので，それぞれの遷移に対し，確率値に基づくスコアを設定する．具体的には，「モーラが撥音である場合には，前のモーラと合わせて 2 モーラで 1 音に対応する確率が高くなる」「改行は長い休符に割り当たりやすい」といった事前に定めたルールを用いて場合を分け，それぞれのパターンごとにスコアを切り替える．スコアを確率値で与えるため，事前に一定数のデータを分析し，各パターンにおけるそれぞれの対応が生起する確率を算出する．その上で，入力全体に対して最大の累積スコアを与える遷移パスを求めることで，対応推定結果を得る．

しかし，従来手法では旋律中の各音から得られる情報として「音符か休符か（音符／休符の違い）」「休符の長さ」のみを使用しており，音符の長さや高さに関する情報を使用していない．したがって，例えば「長い音符は 1 音で 2 モーラに対応しやすい」といった情報が反映できておらず，このような対応が出現する楽曲において推定性能が不十分となっている．

3. 提案する旋律と歌詞の対応推定手法

提案手法では，まず，歌詞中のモーラ・記号を HMM で表現する．モーラは「撥音」「促音」「長音」「それ以外」の 4 種類，記号は「読点」「改行記号」の 2 種類の計 6 種類に分類し，それぞれを後続するモーラ・記号によって場合分けした「バイモーラ」でモデル化する．図 1 にバイモーラ HMM の構造を示す．矢印は入力特徴量に対する遷移を表しており，一番下の経路はこのモーラに音符・休符が割り当たらないことを，

真ん中の経路はこのモーラに 1 音が割り当たることを，一番上の経路は 2 音以上が割り当たることを表している．

入力特徴量は，旋律中の各音符・休符の情報を反映した離散コードで構成されるベクトルとする．具体的には，各音符・休符に対し，「音符／休符の違い（2 種類）」「音符・休符の長さ（4 段階）」「次の音との高さの違い（3 段階）」「次の音との長さの違い（3 段階）」をそれぞれ離散コードで表現して結合し，4 次元の特徴ベクトルを構成する．

i 番目の入力特徴ベクトル $\boldsymbol{o}_i$ に対する HMM の状態 j における出力確率 $b_j(\boldsymbol{o}_i)$ は，以下の式(1)で定義される．このとき，o_{si} は入力特徴ベクトルの s 次元目のコードを表し，$P_{js}(o_{si})$ は状態 j における o_{si} の出力確率，γ_s は各コードに対する重み係数を表している．

$$b_j(\boldsymbol{o}_i) = \prod_{s=1}^{4} P_{js}(o_{si})^{\gamma_s} \qquad (1)$$

対応推定を行う際には，対象となる楽曲の歌詞に従って学習済みのバイモーラ HMM を連結し，1 つの HMM を生成する．そこに，音符・休符情報を反映した特徴ベクトル系列を入力し，Viterbi 探索を行うことで最尤となる状態遷移を求め，その結果からモーラと音符の対応を得る．

4. 評価実験

4.1 実験データ

従来手法における確率値決定のための分析データと，提案手法における HMM の学習データには，J-POP 30 曲と，RWC 研究用音楽データベース[2]のポピュラー音楽 10 曲の計 40 曲を使用した．また，評価データには，RWC データベースのポピュラー音楽 5 曲の A メロとサビ部分，童謡 5 曲のメインパートを使用した．なお，評価データと学習データには重複は存在しない．

4.2 対応推定性能の評価

従来手法と提案手法の対応推定性能の評価結果を図 2 に示す．提案手法については，γ_s を 0 または 1 に設定することで，特徴量のどの次元の情報を利用するかを選択することができるので，いくつかのパターンで評価を行った．図の下の枠内は，音符・休符情報として何が用いられているかを示したものである．なお，評価指標には正解率（すべての対応ペアのうち，正しく推定されたものの割合）を使用した．

結果を見ると，提案手法で「音符／休符の違い」「音符・休符の長さ」を使用した場合，従来手法よりも若干性能が改善している様子が確認できる．これは，「音符の長さ」の情報が加わっていること，従来手法では手動で設定していた対応ルールを HMM の学習により統計的に獲得していることなどに起因していると考えられる．また，「次の音との長さの違い」の情報を加えるとさらに性能が改善し，正解率は 90.7%となった．一方，「次の音との高さの違い」の情報は性能改善には寄与しないことが確認された．

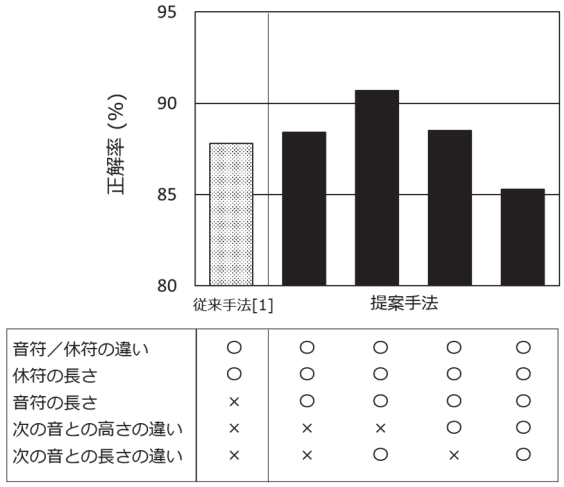

図 2　対応推定性能の評価結果

なお，ポピュラー音楽に対する正解率は従来手法で 95.2%，提案手法で 97.0%，童謡に対する正解率は従来手法で 60.8%，提案手法で 68.0%となった．改善の大きかった童謡の結果を分析したところ，提案手法により「長い音符 1 音に 2 モーラが対応する箇所」が正しく推定され，性能改善につながっていることが確認された．

5. まとめ

本研究では，隠れマルコフモデルを用いた楽曲の旋律と歌詞の自動対応推定手法を提案し，その評価を行った．その結果，提案手法の正解率は最高で約 91%となり，従来手法よりも良好な性能となることが示された．今後は，対応推定のさらなる性能改善や，提案手法の類似度分析システムへの取り込みを図る必要がある．

参考文献

[1] 藤村，岩野，“日本語楽曲の旋律と歌詞のアクセントの関係分析のための自動対応付け,” 情報処理学会第 79 回全国大会講演論文集, vol. 2, 3L-2, pp. 75-76, 2017.

[2] 後藤他，“RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース，” 情報処理学会音楽情報科学研究会研究報告，2001-MUS-42-6, vol.2001, no.103, pp.35-42, 2001.