1N-07

# 歌唱における発声タイミングのずらし時間抽出と発声タイミングモデルの提案

藤田　千尋†　　竹川　佳成‡　　平田　圭二‡

公立はこだて未来大学大学院†　　公立はこだて未来大学‡

## 1. はじめに

近年，VOCALOID 等の音声合成ソフトの普及により個人制作での音声合成ソフトによる歌唱の入った楽曲が増加した．しかし人間らしく歌唱させるにはビブラートやしゃくり等の細かいパラメータ調整が必要になり，特にあえて意図的に発声タイミングを遅らせたり早めたりするような歌い方をさせるには歌唱のリズムを自然になるようにずらして打ち込まなければいけない．本研究の目的は，ずらしを用いた発声タイミングのモデル構築と実際に歌唱へ自動付与するシステムの構築である．実際の歌手による歌唱からずらし表現を用いたフレーズのコーパスを収集し，隠れマルコフモデルによって発声タイミングのモデルを構築する．

## 2. ずらし時間のコーパス生成

ずらし時間の検出方法として，Melodyne によって抽出した歌声旋律を Standard MIDI File(SMF)として抽出し，実際の歌唱での発声タイミングと楽譜通りの歌唱の発声タイミングのオンセット時刻によって比較する．実際の歌唱による SMF と楽譜通りの歌唱による SMF では，子音と母音などの境目やビブラートなどの音高の変化で音符の分割が起こる場合があり，音符数や音高が一部異なってしまうため，単純な一音同士での対応ではオンセット時刻に差が生じてしまう．そこで，動的時間伸縮法(DTW)によって 2 つの譜面同士での距離計算によるマッチングを行う．入力を各音符のオンセット時刻と音高とし，計算によって得られた距離の小さい音符同士が 2 つの歌唱での同一部分として対応する．しかしこの計算では図 1 の(a)のようにしゃくりあげによって発声時の音高が違う場合，予想外に大きなずらしがあった場合，細かい分割があった場合などでは，前後の音符とマッチングしてしまうような誤検出を起こすことがあり，DTW のみでは対応できない．そのため旋律全体のマッチングには DTW を用い，実際の音符の対応と明らかに異なる部分がある場合は歌詞による音符同士のマッチングを行った．図 1 は，平井堅の「君の好きなとこ」の一部のマッチング例である．

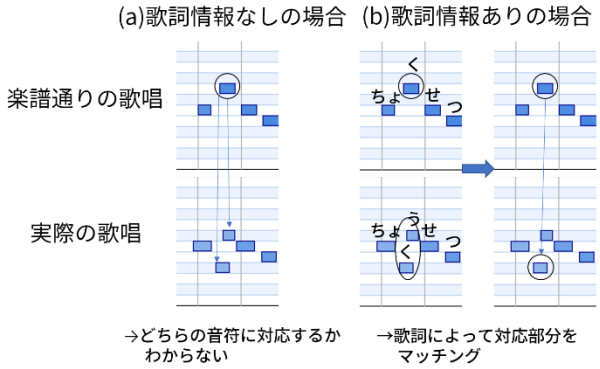


図 1　歌詞によるマッチング

実際の歌唱の SMF と楽譜通りの歌唱の SMF から歌詞情報が一致する音符同士でのオンセット時刻の差を求め，その差をずらし時間とする．SMF からは歌詞情報，ずらし時間，音高情報，オンセット時の拍節情報を抽出し，コーパスとして用いた．現時点では平井堅の楽曲 3 曲から抽出した 1 フレーズにつき約 1 小節の 111 フレーズの譜面情報をコーパスとしている．

## 3. 隠れマルコフモデルによるモデル構築

本研究では平井堅の既存の楽曲の譜面情報を用いて隠れマルコフモデル(HMM)によるずらし時間のモデル構築を行った．図 2 に本研究で用いる HMM の構造を示す．

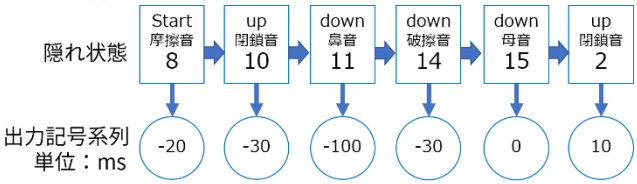


図 2　隠れマルコフモデル

ここでは隠れ状態を上から音高，歌詞，拍節とし，出力記号系列をミリ秒単位のずらし時間とした．音高の遷移としてフレーズの開始点での状態を start，前の音符よりも音階が上がった

Extraction of Shift Time of Phoneme Timing in Singing Voice and Proposal of Model for Producing Phoneme Timing
†Future University Hakodate
‡Future University Hakodate

状態を up，音階が下がった場合を down，音階が同じ場合は same としている．歌詞情報はそのまま扱うと情報量が多くなることに加えて，的場らによる子音長の変化によってグルーブ感に影響が出るという分析[1]がある．そこで，発声方法の違いによってずらし時間の変化に影響するのではないかという仮説を立てて，母音と子音を分類し更に子音の中でも破裂音，摩擦音などのように分類することで計 7 種類に分ける．

前章で抽出した SMF の情報を基に，歌詞，音高，拍節の状態を 1 つにまとめた状態での初期状態確率，状態遷移確率，記号出力確率を算出した．入力された歌声合成ソフトによる譜面にずらし時間を付与するために，学習させた HMM を時間方向に対して前向きに最尤推定し，出力のずらし時間を求めた．

## 4. 評価

図 3，4 に実際の歌唱によるずらし時間と譜面情報に最尤推定を行い導出したずらし時間を示す．図 3，4 で譜面として使用した楽曲は平井堅の「いつか離れる日が来ても」の A メロ冒頭とサビの一部である．また，この楽曲はコーパス未使用である．

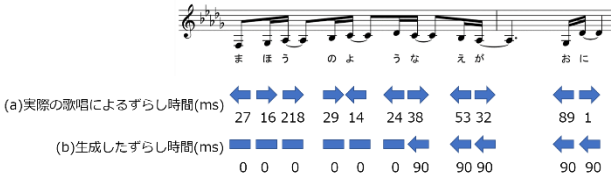


図 3　A メロ冒頭でのずらし時間生成

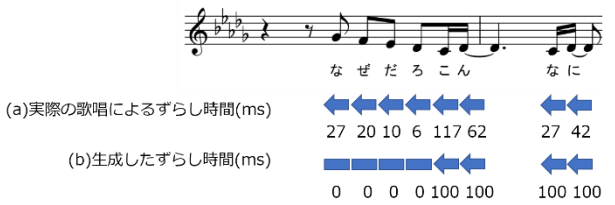


図 4　サビでのずらし時間生成

このように構築したモデルによって生成した発声ずらし時間がコーパスに使用した歌手らしさを与えられているかを評価するために，本研究では平均二乗誤差を使用して評価を行う．評価に使用するものとしてコーパスとして使用した 2 つの楽曲 A，B 中のフレーズ 3 箇所と，コーパス未使用の同歌手による楽曲 C 中のフレーズ 3 箇所にビタビアルゴリズムによる復号を適用し，実際のフレーズでのずらし時間との平均二乗誤差(Mean squared error)を求める．1 フレーズの長さは 2 小節から 3 小節である．以下にずらし時間モデルを適用した各フレーズのずらし時間 $x_i$ (i=0, 1, 2, . . . , n)，実際のフレーズのずらし時間を $y_i$ としたときの平均二乗誤差の式と実際の計算結果を示す．

$$MSE(x) = \frac{1}{n}\sum_{i=0}^{n}(x_i - y_i)^2$$

| | A メロ | 1 番サビ | 2 番サビ |
|---|---|---|---|
| A | 3192 | 8204 | 4764 |
| B | 2238 | 5836 | 5388 |
| C | 5838 | 18084 | 4582 |

表 1　平均二乗誤差の計算結果

表 1 では特に A メロと 1 番サビでコーパス使用曲とコーパス未使用曲に大きな差が見られた．原因としては同歌手による楽曲でも歌詞の構成や旋律の構成が異なるために違いが出てしまったことと，現在のコーパスの多さでは対応しきれない楽曲の構成があることが原因と考えられる．また入力した譜面にコーパス内に存在しない譜面情報を持つ音符が存在する場合，ずらし時間は 1 つ前の音符のずらし時間をそのまま用いるというヒューリスティクスを導入したので，図 3，4(b)のような同じずらし時間が続いたことが結果に影響を与えたのではないかと考える．1 番サビでコーパス使用曲である A，B の数値が大きくなった理由としては，多くの場合 1 番のサビと最後のサビでは同じ歌詞が使用されることがあり，ずらし時間の記号出力確率に影響を与えるためと考えられる．これらの対応として，今後はコーパスの増加とオンセットの拍節情報に楽譜全体での位置情報を与えることを検討し，コーパス内に存在しない譜面情報の際の挙動を見直す必要がある．

## 5. まとめ

本研究の目的はずらしを用いた発声タイミングのモデル構築と，ずらし時間を譜面上の音符に付与するシステムの構築である．現時点でのモデル構築では歌詞情報，音高の遷移，16 分音符単位の発声位置を隠れ状態とした隠れマルコフモデルを利用した結果，楽曲へのずらし時間の適用は出来たが有用な結果を得ることはできなかった．今後はコーパスの増加等を行い，モデルの精度の向上を行う．

## 参考文献

[1] 的場，馬場，成山，松本，森勢，片寄：歌唱のグルーブ感の構成要因の分析；情報処理学会研究報告，Vol.2014-MUS-102，No.12 (2014)．