

市販楽曲からの歌い方ライブラリの作成

池宮 由楽^{1,a)} 糸山 克寿^{1,b)} 奥乃 博^{1,c)}

概要：本稿では、市販楽曲からピブラート、こぶしやグリッサンドといった歌い方に関する特徴を歌唱表現として抽出することで、歌手の歌い方のライブラリを作成する手法について述べる。これらの特徴は、歌唱 F0 軌跡中の特徴的な変動として現れる。本手法ではまず、時間周波数領域での最適経路探索問題を定式化することにより高周波数分解能、高精度な歌唱 F0 推定を行う。推定 F0 軌跡からパターンマッチングにより各歌唱表現を同定、パラメータ表現する。実験では、実際に市販楽曲からプロ歌手の歌唱表現を抽出できることを確認した。

1. はじめに

市販楽曲から歌手の特徴を抽出することにより、歌唱合成や音楽情報検索 [1] などの多くの応用が可能である。本研究では、歌手の“歌い方”の特徴を分析する [2]。

VocaListener [3], Sinsy [4] は共に歌い方を扱う歌唱合成システムである。前者は、ユーザ歌唱のピッチ・音量をボカロイド歌唱へ転写する技術であり、より人間らしい歌い方を実現する。後者は、歌唱音響特徴と楽譜の確率的な関係を学習し、特定歌手の歌声をモデル化する技術である。

本研究は、歌手の歌い方に関わる特徴を歌唱表現として明示的に抽出することで、任意の歌手の歌い方をライブラリ化する技術の開発を目的とする。本稿で扱う歌唱表現は、歌唱 F0 の特徴的な変動成分であるピブラート、こぶし、グリッサンドの 3 種である。分析対象は伴奏付き歌唱とする。

図 1 に本システムの概要を示す。まず、歌唱される音符の音高と順序を表す音高列による制約を用いた歌唱 F0 推定を行う。歌唱 F0 からテンプレートに基づき歌唱表現を同定し、各表現をよく表すパラメータとして保存することでライブラリを構築する。また、応用としてプロ歌手の歌い方を合成歌唱へ転写するシステムを作成した。

2. 伴奏付き歌唱からの F0 推定

F0 推定は、入力音響信号の対数周波数スペクトログラム上での、最も歌唱 F0 らしい時系列の探索により実現する。スペクトログラムは定 Q 変換で得る。時系列探索は歌唱 F0 軌跡の滑らかさを考慮したマルコフモデルとして定式化し動的計画法により解く [2]。また、F0 推定誤り抑制のため入

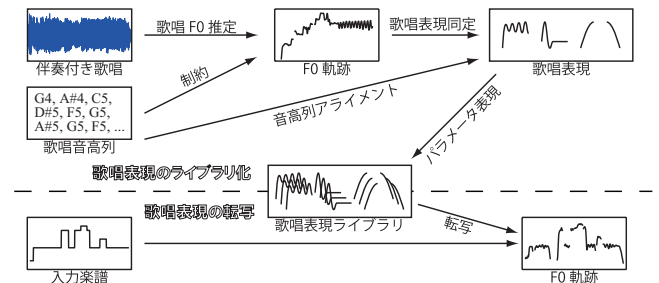


図 1: 提案システムの概要

力音高列に基づいて周波数の探索範囲を制限する。

3. 歌唱表現のライブラリ化

まず歌唱 F0 と音高列を二乗誤差最小化に基づきアライメントし、各音高について個別に処理が行われる。F0 軌跡から抽出した特徴点からパラメータを計算し、それらが 3.2 節で述べるテンプレートに当てはまるとき、各歌唱表現として同定する。図 2 に各歌唱表現のパラメータ表現を示す。

3.1 歌唱 F0 軌跡の特徴点抽出

歌唱 F0 軌跡に含まれる極値点と折れ点を特徴点として抽出する。ここで、折れ点は値が急激に変化する点であり、具体的には、 $|\Delta\Delta F0|$ が $500 [\text{cent} / (\text{sec})^2]$ 以上の点とする。

3.2 歌唱表現のパラメータ表現

[ピブラート] 3 つ以上連続した特徴点について、 $i (> 1)$ 番目の特徴点の値と時間を f_i, t_i とし、対応する周波数 R_i [Hz] と振幅 E_i [cent] を以下の式で定義する。

$$R_i = \frac{1}{t_{i+1} - t_{i-1}} \quad \text{and} \\ E_i = |(f_{i+1} - f_{i-1})(t_i - t_{i-1})R_i + (f_{i-1} - f_i)|.$$

周波数が 3 [Hz] 以上、且つ振幅が 30 [cent] 以上である特徴点をピーク点と定義する。ピーク点が 4 つ (2 周期分) 以

¹ 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

a) ikemiya@kuis.kyoto-u.ac.jp
b) itoyama@kuis.kyoto-u.ac.jp
c) okuno@kuis.kyoto-u.ac.jp

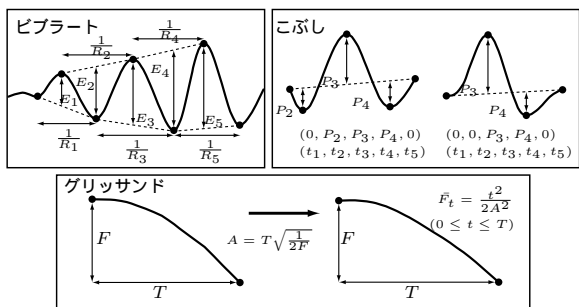


図 2: 歌唱表現のパラメータ表現

上連続する区間をビブラートとして同定する，保存パラメータは $(\{R_n, E_n\}_{1 \leq n \leq N})$ となる．ここで， N はビブラートに含まれるピーク点の個数である．

[グリッサンド] グリスダウンは，フレーズ終りの音符の最後尾における，落ち幅が F_{least} [cent] 以上の単調減少として同定される．予備実験の結果に基づき， F_{least} は 200 とする．グリスダウンの形を，始点に極大値を持つ放物線としてモデル化し，放物線の係数 A と時間幅 T [sec] をパラメータとする．ここで，グリスダウンの落ち幅を F [cent] とすると， $A = T\sqrt{1/(2F)}$ である．

[こぶし] ビブラート以外の区間について，ピーク点を検出する．このとき，振幅が 150 [cent] 以上のピーク点をメインピークとし，その前後にピークがあればサブピークとし，それらに始点，終点を含んだ区間をこぶしとする．サブピークが存在しない場合，始点または終点と同じ特徴点とする．これより，こぶしのパラメータ表現は 5 要素のリストとなる．各要素は，こぶしの始点，左サブピーク，メインピーク，右サブピーク，終点におけるピーク値と時間の組である． i 番目の要素のピーク値 P_i については，以下の式で計算される．

$$P_i = f_i - \left(\frac{f_5 - f_1}{t_5 - t_1} (t_i - t_1) + f_1 \right),$$

ここで， t_i と f_i はそれぞれ i 番目の特徴点の時間と対数周波数である．

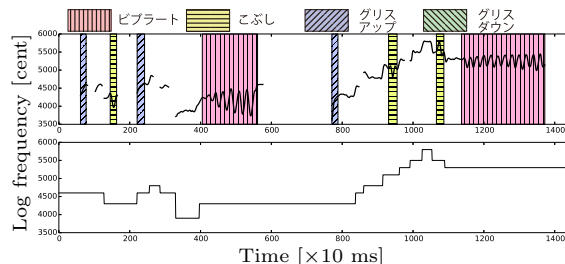
4. 評価実験

4.1 実験条件

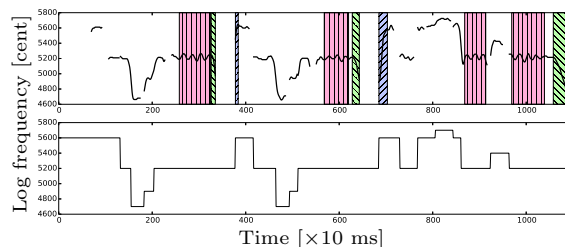
全ての楽曲はサンプリング周波数 16 kHz，量子化数 16 bit である．無歌唱区間は事前に検出されているとする．

4.2 市販楽曲からの歌唱表現の同定と再合成

提案手法を市販楽曲，“人生一路（美空ひばり）”の A メロ部と“クリスピー（スピッツ）”のサビ部に適用した結果を Fig. 3 に示す．前者では，大きなビブラートやこぶしといった演歌に特徴的な歌唱表現が同定されている．後者では，ロングトーンでのグリスダウが同定されており，これはスピッツの歌い方をよく特徴付けているものである．また，図 3(a) におけるこぶしと，3(b) におけるグリスダウを再合成したところ，二乗平均平方根誤差はそれぞれ 16.0，22.3 [cent] となった．100 [cent] が半音であるから，忠実な再合成が行えていると言える．



(a) 人生一路（美空ひばり）



(b) クリスピー（スピッツ）

図 3: 歌唱表現のライブラリ化．上側の図は推定 F0 と同定され歌唱表現を，下側の図は音高列と F0 とのアライメントを示す．

4.3 歌唱表現の転写

美空ひばりとスピッツの楽曲（それぞれ 6 ピース）から作成した歌唱表現ライブラリから，二種の歌唱合成器に対して，歌唱表現の転写を行った．用いた合成器は Vocaloid と Cevio である．Vocaloid ではピッチベンドにより各時刻の音符音高からの差分を操作可能で，Cevio では各時刻の対数 F0 を直接操作可能である．各ライブラリから転写された歌唱を聴取したところ，いずれの合成器においても，無転写のものと比較し各歌手らしさを備えた歌唱となっていることを確認した．各歌唱は我々のサイトで聴くことができる*1．

5. おわりに

本稿では，歌い方の特徴として歌唱表現を保存することでライブラリを作成する手法を提案した．評価実験では，市販楽曲から歌唱表現を抽出できることを確認し，合成歌唱へ歌い方を転写するシステムを構築した．今後は，ライブラリ化や転写の定量評価や，歌い方に基づく検索への応用などを行っていく予定である．なお，本研究は科研費 (S) No. 24220006 の支援を受けた．

参考文献

- [1] Downie, J.S.: “Music information retrieval.”, *Annu. Rev. Inf. Sci. Technol.* 37, pp. 295–340, 2003.
- [2] 池宮 由梁, 糸山 克寿, 奥乃 博: “伴奏付き歌唱に含まれる歌い方要素の個別抽出”, 音楽情報科学研究会, No. 20, pp.1–6, 2013.
- [3] T.Nakano and M. Goto: “VocaListener: A Singing-to-Singing Synthesis System Based on Iterative Parameter Estimation”, *Proc. SMC*, pp.343–348, Sep. 2009.
- [4] K. Oura et al., eds: “Recent Development of the HMM-based Singing Voice Synthesis System - Sinsy”, *SSW7*, pp.211–216, 2010.

*1 winnie.kuis.kyoto-u.ac.jp/members/ikemiya/demo/sst2013.html