5R-7

伴奏付き歌唱からの歌唱表現のパラメータ化と転写

池宮 由楽‡　糸山 克寿‡　奥乃 博‡
‡京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

伴奏付き歌唱からの歌手特徴の抽出は，歌唱合成や音楽情報検索 [1] などの多くの応用で不可欠な技術である．本研究では，歌手の “歌い方” の特徴を分析する [2]．歌い方の特徴を抽出し，合成歌唱に転写することで，その歌い方を再現する．

VocaListener[3] と 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，ユーザ歌唱の歌い方をボーカロイドの歌唱へ転写する技術であるが，歌い方の特徴の解析は行っていない．後者は，歌唱の音響特徴と楽譜の関係を確率的に学習し歌唱を合成する技術であるが，学習に単一歌唱と楽譜の大量のセットが必要であり，また，歌い方の特徴を明示的に扱うことができない．

本稿では，歌手の歌い方に関わる音響特徴を歌唱表現として分析抽出，ライブラリ化し，その歌い方を合成歌唱に転写するシステムを報告する．本稿で扱う歌唱表現は，歌唱 F0 に含まれる特徴的な変動成分であるビブラート，こぶし，グリッサンドの 3 種である．分析対象は，市販楽曲の伴奏付き歌唱である．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．歌唱表現の転写はルールベースで，パラメータから再合成された歌唱表現を入力楽譜中の各音符に付加することで行う．

2. 伴奏付き歌唱からの F0 推定

F0 推定は，入力音響信号から対数周波数スペクトログラムを計算し，最も歌唱 F0 らしい時系列を探索することによって行う．このとき，入力音高列から探索する周波数範囲を制約する．計算には定 Q 変換を用い，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].

3. 歌唱表現のライブラリ化

歌唱 F0 軌跡から歌唱表現を同定し，保存する．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．このとき，パラメータとその音符情報をまとめて一つの要素とすることでインデクシングを行う．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立ち上がり（下がり）点を特徴点として抽出する，ここで，立ち上がり（下がり）点は値が急激に変化し始める（終わる）点を表している．具体的には，立ち上がり点は $\Delta\Delta F0$ が 500 [cent / (sec)^2] 以上の点として抽出する．

3.2 歌唱表現のパラメータ表現

[**ビブラート**]　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

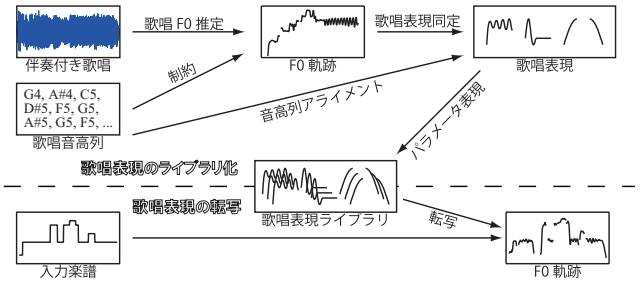


図 1: 提案システムの概要

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）以上連続する区間をビブラートとして同定する，音符内でのビブラート開始位置 S [sec] とすると，ビブラートパラメータは $(S, \{R_n, E_n\}_{1\le n\le N})$ 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．グリスアップはフレーズ始まりの先頭における単調増加であり，グリスダウンを左右反転させた定義を用いる．

[**こぶし**]　他の歌唱表現ではない区間について，ビブラートと同様に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，いかの式で計算される．

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

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．

3.3 音符情報

各歌唱表現が同定された音符の情報を付随情報として保存する．本稿で扱う音符情報は，音高（ノートナンバー），音長，前後の音符との音高差と音符ラベルである．ここで音符ラベルは，音符がフレーズ始まり，途中，終りのいずれであるかを示したラベルであるとする．

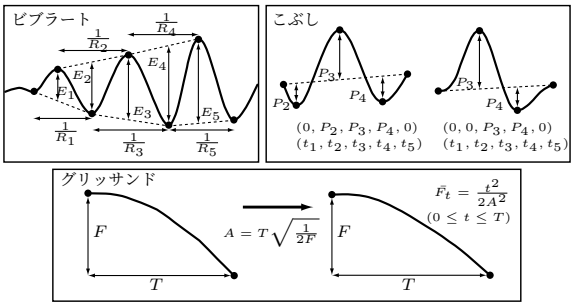


図 2: 歌唱表現のパラメータ表現

4. 歌唱表現の転写

本章では，歌唱表現ライブラリを用いて，歌唱合成器による歌唱に特定歌手の歌い方を転写する手法について述べる．具体的には，テンポ・音符長を含む楽譜情報が与えられたとき，ライブラリの音符情報から適切な歌唱表現を選択し，各音符に付加する．ここで対象とする歌唱合成器は，局所的にピッチを変動することを可能とする何らかの機構を有するものとする．

4.1 前処理

入力楽譜と歌唱表現ライブラリの音域が大きく違う場合，歌唱表現を正しく転写できないと考えられる．そのため，楽譜とライブラリの最低音高が合うように，音高をシフトする．また，楽譜中の全音符について，ライブラリと同様の音符情報を取得する．

4.2 転写ルール

楽譜中の各音符（目標音符とする）について続く処理を行う．まず，ライブラリから次の 4 条件に適合する全ての歌唱表現パラメータを取得する．

1. **(全歌唱表現)** 音符ラベルが目標音符と等しい．
2. **(全歌唱表現)** 目標音符との音高の差が M 以下．
3. **(こぶし・グリッサンド)** 時間長が目標音符長よりも短い．
4. **(こぶし)** 前後の音符との音高差の符号が目標音符のものと等しい．

M が小さいほど，よりルールが厳しくなると言える．

次に各歌唱表現について，取得されたパラメータのうち，音符情報が目標音符に最も近いものを選択し，付加する．ここで，パラメータが一つも取得されていない場合，その歌唱表現は付加されない．音符情報の近さの指標と優先度は，1: 音高の差，2: 音長の差，とする．歌唱表現はパラメータから再合成され，目標音符の F0 軌跡上に貼り付けることにより転写される．

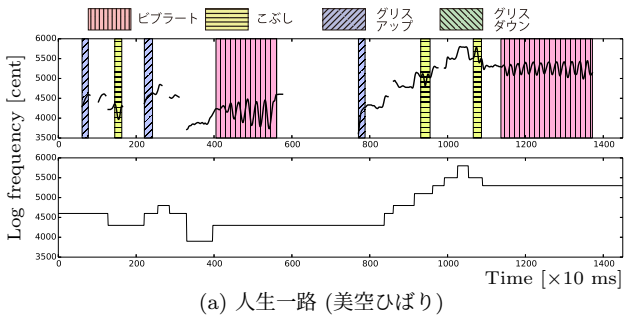
5. 評価実験

5.1 実験条件

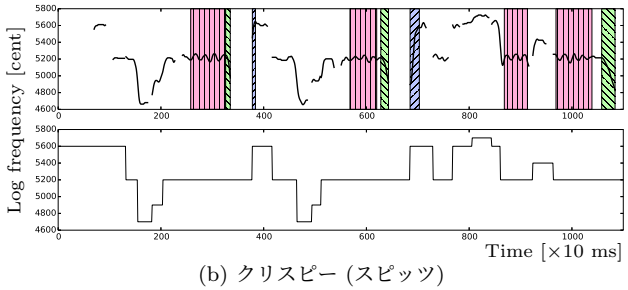
実験に用いる全ての楽曲はサンプリング周波数 16 kHz，量子化数 16 bit である．また，無歌唱区間は事前に検出されているとする．

5.2 市販楽曲からの歌唱表現の保存

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ンが同定されており，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ン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，提案したパラメータ表現は十分正確な歌唱表現の再合成が可能であると言える．



(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．

5.3 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§．

6. 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し，合成歌唱へ歌い方を転写するシステムを提案した．評価実験では，市販楽曲からライブラリを作成し，二種の合成歌唱に対し転写を行うことで，本システムの有効性を確認した．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 由楽, 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, A. Mase, T. Yamada, S. Muto, Y. Nankaku and K. Tokuda: “Recent Development of the HMM-based Singing Voice Synthesis System - Sinsy”, *Proc. ISCA Tutorial and Research Workshop on Speech Synthesis*, pp.211–216, 2010.

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