1T-03

同一楽曲に対する多数の歌唱の音高推移分布および再生数の可視化

近藤　芽衣†　伊藤　貴之†　中野　倫靖‡　深山　覚‡　濱崎　雅弘‡　後藤　真孝‡
お茶の水女子大学†　産業技術総合研究所‡

1　はじめに

2 次創作やソーシャルメディア環境の普及にともない，自らの歌唱を録音・録画して動画共有サービスへ投稿する機会が増えた．その結果，同一楽曲を様々な歌唱者が歌った音源を，人々が鑑賞して楽しめるようになった現状がある．

このような歌唱群に対し，個々の歌唱者の癖や個性を理解するための一手段として我々は，それらの歌唱音響データからそれぞれの音高の推移を抽出し，その分布を可視化する手法を開発している．そのような手法の一つとして，本稿著者でもある伊藤らが提案した SingDistVis [1]は，音高および時刻を 2 軸とするヒストグラムによる可視化と，その局所部分をズームアップした折れ線表示での可視化により，音高の特徴的な分布の発見を支援する．折れ線表示には，サンプリングで同時に表示する本数を制御することで Visual Cluttering を防いでいる．

本報告では SingDistVis の応用として，ソーシャルメディア上の再生数で各歌唱を色分けし，音高分布との関係を可視化した事例を報告する．

2　処理手順

本章では提案手法の各処理を手順に沿って示す．

2.1　音楽音響信号からの歌声の音高推定

多様な音源を可視化対象とするために，伴奏がミックスされた歌唱音源（混合音）を扱う．まず混合音から歌声のみを分離し，その分離された歌唱音源から音高を推定する 2 段階の処理を行うが，それぞれ以下の手法を用いた．

2.1.1　Spleeter による歌声分離

U-Net 構造を持つ深層学習ベースの音源分離手法である Spleeter [2]を用いて，混合音から歌声を分離する．入力音響データはステレオ MP3 形式，サンプリングレートは 44100Hz とする．

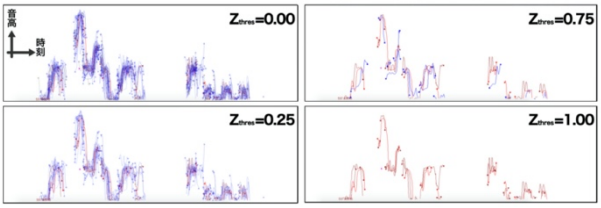

図 1　サンプリングによる折れ線の本数制御

2.1.2　基本周波数（F0）推定

Spleeter により分離した歌声から，音高として基本周波数（F0）を推定する．混合音から伴奏音を完全に除いて歌声分離するのは難しいため，F0 推定手法は耐雑音性に優れたものが望ましい．そこで本研究では PYIN [3]を用いた．

2.2　SingDistVis による音高可視化

音高の可視化について，SingDistVis の処理手順を概説する．詳細は文献 [1]を参考にされたい．

2.2.1　音高データの表記

本章では歌唱集合Sを構成する各歌唱の音高推移を以下のように表記する．

$$S=\{s_1, s_2, \cdots, s_i, \cdots s_I\}$$
$$s_i = \{e_i, p_{i1}, p_{i2}, \cdots p_{ij}, \cdots p_{iJ}\}$$

ここでs_iはi番目の歌唱の音高系列，Iは歌唱者の総数，e_iはi番目の歌唱者の歌唱動画の再生数に応じた評価係数である．p_{ij}はi番目の歌唱者のj番目の時刻における F0 値の対数，Jは F0 推定の対象区間における標本化された時刻の総数（F0 値の個数）である．なお休符に相当する無音部分には，便宜上，F0 値の対数にゼロを代入した．

現状の実装では評価係数e_iは 4 段階となっており，最も再生数が低い歌唱群に$e_i = 1$を，最も再生数が高い歌唱群に$e_i = 4$を付与する．なお，原曲にあたる歌唱には，他の歌唱と区別するために$e_i = 5$を付与する．楽曲により評価係数を定める再生数の閾値は異なり，実行例では 5000 回再生以上の歌唱に$e_i = 4$を付与している．

2.2.2　ヒストグラム画像の生成

本手法では，時刻を横軸，音高を縦軸とした長方形領域を設定し，これを格子状に分割する．p_{ij}の各々が上述の格子構造のいずれの長方形領域に該当するかを算出し，各格子領域を通過した歌唱数から各格子領域の濃淡を算出することで，グレースケールのヒストグラム画像を生成する．

Visualization of the distribution of pitch transition and the number of views of many singings for the same song.
† Mei Kondo, Ochanomizu University
† Takayuki Itoh, Ochanomizu University
‡ Tomoyasu Nakano, AIST
‡ Satoru Fukayama, AIST
‡ Masahiro Hamasaki, AIST
‡ Masataka Goto, AIST

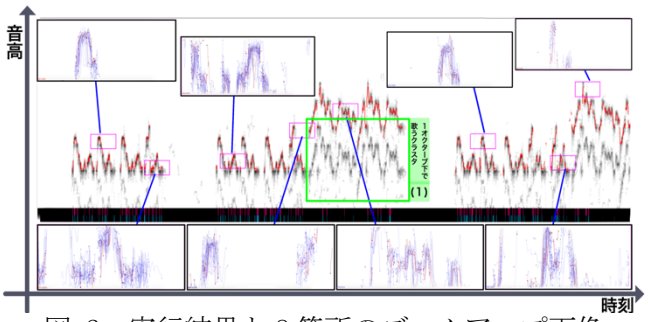

図 2　実行結果と 8 箇所のズームアップ画像

2.3　SingDistVis の拡張

SingDistVis の GUI におけるヒストグラム画像において，ユーザが指定した矩形領域に対応する音高推移を，折れ線の集合で表現する．この際 Visual Cluttering を防ぐために，同時に描画する折れ線の本数をサンプリングにより制御する．これは，各折れ線p_i.に対して，ユーザ指定のタイミングで再計算できる．一様乱数z_i $(0.0 \leq z_i \leq 1.0)$ を割り当て，以下を満たす場合のみ描画する．

$$\beta_{e_i} z_i > Z_{thres}$$

ここで，β_{e_i}は評価係数e_iに応じた係数，Z_{thres}は GUI スライダーで調整する閾値であり，いずれもユーザが調節可能なパラメータである．

図 1 は折れ線制御の図である．原曲にあたる折れ線はZ_{thres} がどの値をとっても消えることはない．また，本数を少なく設定すると，再生数の高い音源の折れ線のほうが画面に残りやすいように選択表示する．

3　実行例

本手法による可視化の例を紹介する．プログラミング環境は Java 1.12.0 および JOGL (Java binding for OpenGL) 2.3.2 を用いた．実行例には【初音ミク】　夜明けと蛍　【オリジナル】（https://www.nicovideo.jp/watch/sm24892241）の 67 人の歌唱を用い，再生数はニコニコ動画における再生数を採用した．本報告では音響データから Spleeter [2]の 2stem モデルを用いて音源を分離し，PYIN [3]を用いて推定した F0 を入力とした．可視化結果の画素数は$N = 1000, M = 480$とし，対象となる周波数を 110Hz～1760Hz（オクターブ表記付き音名で A1 から A5）の 4 オクターブとした．

図 2 は音高推移分布をヒストグラム画像として表示した例とその中で 8 箇所をズームアップし表示した例である．グレースケールで黒に近い箇所では，同じような音高推移の歌唱が多いことを意味する．例えば，図中の緑枠（1）では音高推移が二つに分かれていることが分かるが，各遷移が同じ濃さであることから，サビに入ってから音高を 1 オクターブ低く歌唱した人も多かったことがわかる．

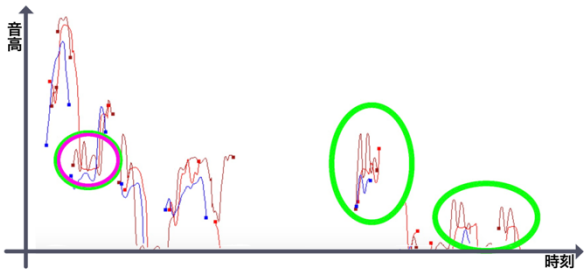

図 3　サビ冒頭のズームアップ画像

図 3 では赤色の折れ線が再生数の高い歌唱，青色の折れ線が再生数の低い歌唱を示している．

特に赤色の歌唱について，ビブラート（周期的に音高を変動させる歌唱テクニック）とみられる音高の揺れを確認することができた．

4　まとめ

本報告では，多数の歌唱者による同一楽曲に対する伴奏付きの歌唱音源の音高推移分布と，それぞれの再生数を可視化した例を示した．多くの歌唱が原曲とは異なる音高推移を描いているケースや，サビでオクターブ下げて歌う歌唱者が複数見られるケースのほか，再生数の高い歌声と低い歌声では，ビブラートの有無など音高の変化に違いが見られた．

今後の展望として，ビブラートなどのボーカルテクニックを特徴量として抽出し，可視化画面への表示を加えることで，より詳細な分析を可能としたい．また異なるキーで歌唱された同一楽曲について，キーを合わせて可視化する機能を追加予定である．また GUI 機能の拡充として，可視化画面からの音源再生機能や，ユーザ自身の歌唱を区別して表示する機能を追加することで，好きな歌唱を選んでそれを目標として歌唱を練習するための支援ツールを開発したい．

参考文献

[1] 伊藤 貴之, 中野 倫靖, 深山 覚, 濱崎 雅弘, 後藤 真孝, SingDistVis:多数の歌声から歌い方の傾向を可視化できるインタフェース, ソフトウェア科学会 WISS 2021 論文集, 94, pp.1-8, 2021.

[2] R. Hennequin, A. Khlif, F. Voituret, M. Moussallam, Spleeter: A Fast and Efficient Music Source Separation Tool With Pre-trained Models, Journal of Open Source Software, 5(50):2154, 2020. doi: 10.21105/joss.02154.

[3] M. Mauch, S. Dixon, PYIN: A fundamental frequency estimator using probabilistic threshold distributions, Proc. ICASSP 2014, pp. 659-663, 2014.