同一楽曲に対する多数の歌唱と正解歌唱の音高推移分布の可視化

近藤 芽衣[1,a)]　伊藤 貴之[1,b)]　中野 倫靖[2,c)]　深山 覚[2,d)]　濱崎 雅弘[2,e)]　後藤 真孝[2,f)]

概要：同一楽曲に対して多数の歌唱者がソーシャルメディアなどに自分の歌唱作品を公開する機会が近年増えている．このような歌唱群の癖や個性を理解するための一手段として我々は，同一楽曲に対する歌唱者群の歌い方を可視化する手法を開発している．この手法では，同一楽曲に対する多数の歌唱群の音響データからそれぞれの音高（基本周波数：以下 F0 と称する）の推移を抽出し，その分布を可視化する．また視覚表現の手段として，時刻および周波数の対数値を 2 軸とする 2 次元ヒストグラム画像を生成し，これに適応 2 値化・ラベリングといった画像処理手法を適用している．本報告ではこの可視化手法と可視化結果を紹介したのちに，一般的な歌唱音響データから無伴奏歌唱音響データを生成する手順，およびその結果として得られた音高推移分布の可視化結果を示す．特に本稿では，原曲での音高推移を正解音高推移として可視化する機能を追加開発した結果を新しく示す．

1. はじめに

多数の歌唱者が同一の楽曲に対する歌声を公開する機会が近年増えている．このような多数の歌唱者の歌唱データを分析することは，学術的な観点からも興味深く，また歌唱力向上や流行分析などの実用の可能性も秘めている．例えば，どのような癖を有する歌唱者が人気を博する傾向にあるか，楽譜上の音高に対しどのようにバリエーションのある表現をする歌唱者がいるかを分析できる．また，プロの歌唱者と他の歌唱者の歌唱を比較することで，プロの歌唱者のどのような歌唱の模倣が容易か困難か，といった議論も可能になる．このような歌唱データ分析において，本稿では楽曲に対しどのようにバリエーションのある表現をする歌唱者がいるのかを分析することを対象とする．

歌唱データにおける音高の推移は一種の時系列データとして捉えることができる．時系列データの分類や特徴検出には従来から多数の研究がなされており [1], [2]，それらを適用することで歌唱の癖の違いや音高の逸脱を検出することが可能である．一方で，歌唱の分析においては音高の逸脱を発見するだけでなく，その意味付けが重要である場合もある．例えば同一楽曲の特定の瞬間の音高に個人差が見られた際に，意図的な歌唱技法として音高をずらしているのか，技量不足や練習不足により意図せずに音高がずれているのか，といった点を解釈するには分析者の理解や判断が必要である．そこで我々は，主観的・定性的に歌唱データを観察するための可視化手法を開発している．可視化は大規模データを理解するために多用されており，多数の歌唱データを理解することにより多数の歌唱者の歌い方のバリエーションを分析するという本研究の目的にも合致している．時系列データに対する汎用的な可視化手法は従来から多数研究されており，これを歌唱データの分析に適用することも可能である．

我々は同一楽曲に対する多数の歌唱の音高の推移分布を可視化する一手法を提案している [3]．本手法では一定時刻ごとに推定した F0 を時系列情報とみなし，横軸を時刻・縦軸を F0 の対数値とする格子の上にプロットする．格子上の各長方形領域におけるプロット回数を集計することで，2 次元ヒストグラムを構成する．この 2 次元ヒストグラムをグレースケールの画像として表示し，一定以上の明度を有する連続領域をラベリングすることで，多くの歌唱者が同様に用いる音高の推移を可視化する．

本報告では我々の可視化手法および可視化例を紹介するとともに，一般的な歌唱音響データから無伴奏歌唱音響データを生成する手法と，この手法によって得られた音高推移分布の可視化例を示す．特に本稿では，原曲での音高

[1] お茶の水女子大学
Ocha University
[2] 産業技術総合研究所
AIST
[a)] g1620517@is.ocha.ac.jp
[b)] itot@is.ocha.ac.jp
[c)] t.nakano@aist.go.jp
[d)] s.fukayama@aist.go.jp
[e)] masahiro.hamasaki@aist.go.jp
[f)] m.goto@aist.go.jp

推移を正解音高推移として可視化する機能を追加開発した結果を新しく示す．

2. 関連研究

2.1 多数の歌唱データの活用

同一楽曲に対する多数の歌唱データがあれば，歌唱技法の傾向を分析することが可能になる．この点に着目した研究の例として Wilkins ら [4] は，20 人のプロ歌唱者による 10 時間以上の歌唱録音データベースを構築し，ビブラートやトリルといった歌唱技法を分析した結果を示している．

また歌唱作品の制作環境の一例として Tsuzuki ら [5], [6] は，同一楽曲に対する複数歌唱を組み合わせて合唱作品を制作する過程を支援するツールを提案している．

2.2 音高データの分析

歌唱の癖や違いを測る基本的な指標の一つに音高（F0）があげられる．歌唱データの各時刻における音高を推定し，楽譜から得られる音高と比較することで，音高をあえて外す意図的な歌唱，あるいは音高が大きく逸脱している歌唱などを発見できる [7]．あるいはビブラート等の歌唱テクニックを F0 の推移から検出したり [8], [9]，オーバーシュートなどの動的変動 [10] に着目したりすることなどにより，歌唱の個性を分析できる．

2.3 音高推移の可視化

歌唱データおよびそれに限定しない演奏データにおける音高推移の分析や観察に可視化を用いた研究事例は，既にいくつか報告されている [11]．例として，Nakano ら [12] の MiruSinger，Shiraishi ら [13] の HAMOKARA，Moschos ら [14] の FONASKEIN，Mayor ら [15] の歌唱採点手法では，歌唱の練習成果を正解楽譜（ピアノロールなど）と比較可視化する機能を搭載している．複数の歌唱を対象とした例として，Nakano ら [16] の VocaRefiner は複数の歌唱録音の編集により楽曲を制作するための対話的環境を構築しており，この中で音高の可視化も採用している．

また演奏情報の中から基本周波数およびその時間推移の適切な同定を支援するための可視化 [17], [18] や，周波数情報から推察される調性の可視化 [19] などの事例がある．また，歌唱の音高推移から歌唱スタイルを理解するために可視化を用いる手法として，音高とダイナミクスを 2 軸とした可視化 [20] や，音高と音高差分を 2 軸とした可視化 [21] が報告されている．Wilkins[4] らによる歌唱技法の同定結果はスペクトログラムとして可視化されている．

しかし我々が調査する限り，数百人・数千人単位の多数の歌唱データを一画面に可視化する研究事例は見当たらない．

2.4 時系列データの可視化

歌唱の音高の推移は時系列データとして扱うことが可能であり，汎用的な時系列データ可視化手法を適用することが可能である．

ここで n 個の標本がそれぞれ m 個の時刻における実数値を有する時系列データがあるとする．このようなデータに関する多くの可視化手法は以下のいずれかのアプローチを有する．なお，以下での「実数値」は F0 値に，「密度」は近い音高を有する歌唱者の人数に対応する．

(1) 一方の座標軸に m 個の時刻，他方の座標軸に実数値を割り当てた折れ線グラフ [22], [23] や散布図．
(2) (1) の折れ線や点群を密度関数に置き換えて，密度を各画素の明度や色相に変換したヒートマップで表現したもの [24]．
(3) 一方の座標軸に m 個の時刻，他方の座標軸に n 個の標本を割り当てたマトリクスに対して，実数値を各画素の明度や色相に変換したヒートマップで表現したもの [25], [26]．

これらのアプローチの各々にはいくつかの問題点がある．(1) に示した折れ線グラフや散布図には，画面上の描画物の過密状態が引き起こす Visual Cluttering と呼ばれる視認性の低下が避けられない．また可視化結果からのデータ読み取りにおいて色の識別能力は高くない [27] ことが知られており，(3) に示したヒートマップでは実数値を正確に読み取れない可能性がある．以上により本研究では (2) に示す「密度関数のヒートマップ」というアプローチをとることにする．

本研究の目的の一つとして「歌唱における音高推移のパターンを発見する」という点がある．時系列データの可視化においても，クラスタリングや部分頻出パターン検出などの汎用的な手法を用いている事例がいくつかある [28]．ここで音高推移の部分パターンには「同一の瞬間にも異なる時間長のパターンが同時に出現しうる」という特徴があり，この点に着目した時系列データ可視化手法はまだ多くない．本研究が採用する「密度関数のヒートマップ」は画像の一種であり．このような画像に領域分割手法を適用することで，密度関数のヒートマップから密度の濃い領域を抽出し，これをクラスタとみなすことが可能である．本研究ではこのアプローチによって，歌唱における音高推移のパターンを可視化する．

3. 音高分布の可視化

本章では，我々が提案している音高（F0）の可視化のための画像処理的なアプローチ [3] について，処理手順に沿って論じる．

3.1 音高データの表記

本章では歌唱者集合 S を構成する各歌唱者の音高の推移

を以下のように表記する．

$$S = \{s_1, s_2, ..., s_n\}$$
$$s_i = \{p_{i1}, p_{i2}, ...p_{im}\} \quad (1)$$

ここで s_i は i 番目の歌唱者による歌唱の音高系列，n は歌唱者の総数，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，m は基本周波数推定の対象区間における標本化された時刻の総数（各音高系列の 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．

なお本研究では，全ての歌唱が同じ長さ・同じタイミングで収録された上で，同一の時刻における F0 を推定することを前提としている．

3.2 グレースケール画像の生成

本手法では，時刻を横軸，周波数の対数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．歌唱の開始時刻および終了時刻をそれぞれ $t_{\mathrm{start}}, t_{\mathrm{end}}$ として長方形領域 R の左右端にわりあて，この区間を N 個に分割する．また可視化の対象となる周波数領域の上限と下限を設定し，各々の対数をそれぞれ $p_{\max}, p_{\min}$ として R の上下端にわりあて，これを M 個に分割する．なお，以下の記述では $t_{\mathrm{start}} = t_1, t_{\mathrm{end}} = t_N, p_{\min} = f_1, p_{\max} = f_M$ とする．

続いて本手法では，式 1 に示す 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する．具体的には，左から u 番目，下から v 番目の長方形領域について，

$$t_u < i < t_{u+1}$$
$$f_v < p_{ij} < f_{v+1} \quad (2)$$

が成立するようであれば，p_{ij} は当該長方形領域に該当するとして，変数 r_{uv} に 1 を加算する．

以上の処理による集計結果は 2 次元ヒストグラムを構成するが，本手法ではこれを横 N 画素，縦 M 画素の画像として扱う．長方形領域に包括される p_{ij} の個数を集計した変数 r_{uv} から，以下の式

$$I_{uv} = 1.0 - (\alpha r_{uv})^{\gamma} \quad (3)$$

によって，左から u 画素目，下から v 画素目の明度 I_{uv} を求める．ここでの α および γ はユーザが調節可能な変数とする．

3.3 ラベリングによる頻出周波数推移領域の特定

頻出する周波数推移を見つけやすくするための一手段として，本手法では上述の画像を閾値 β によって白黒 2 値化し，さらにラベリング処理を適用する．まず以下の式

$$B_{uv} = 1(I_{uv} > \beta)$$
$$B_{uv} = 0(I_{uv} \leq \beta) \quad (4)$$

によって，左から u 画素目，下から v 画素目の画素値 I_{uv} を 1 または 0 のいずれかを有する画素値 B_{uv} に変換する．続いて $B_{uv} = 1$ である画素を 1 個抽出し，隣接画素で $B_{uv} = 1$ であるものを再帰的に探索する．そして，探索が終了するまでに訪問した画素の集合に固有のラベルを割り当てる．この処理を $B_{uv} = 1$ である全ての画素に割り当てることで，一定以上の頻度で現れる周波数推移領域を抽出する．なおラベリング結果は画素の処理順に依存しない．閾値 β はユーザが調節可能な変数である．

3.4 可視化の例

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 1.10.0 および JOGL (Java OpenGL) 2.3.2 を用い，実行例には DAMP-balanced dataset [*1] に収録された “Let It Go” の 2024 人の歌唱を用いた．DAMP-balanced dataset には F0 値の推定結果を記述したデータファイルも収録されているが，本報告では我々自身で音響データから STRAIGHT[29] 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 $N = 1000, M = 500$ とした．

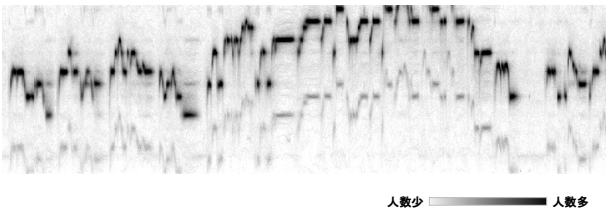

図 1 音高推移分布をグレースケール画像として表示した例．黒に近いほど多くの歌唱者が同じ音高推移をとっていることを示している．

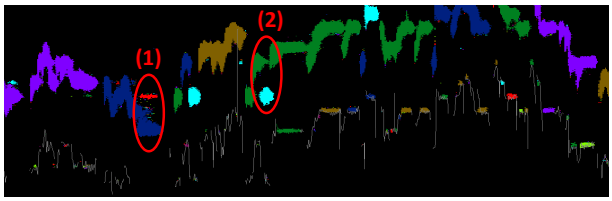

図 2 グレースケール画像を 2 値化してラベリングした例．ラベルごとに異なる色が割り当てられている．ただし現時点での実装では所定の 10 色のいずれかがラベルに付与されているため，複数のラベルに同一色が割り当てられている箇所がある．

図 1 は周波数推移分布をグレースケール画像として表示した例である．画像中の上部に黒に近い部位が左右に分布しており，多くの歌唱者が同様な音高をとっていることがわかる．さらに，これと同様な動きが画像中の下部にも薄く見える．このような部位が見られる理由として，少数の

[*1] https://ccrma.stanford.edu/damp/

歌唱者が音高を 1 オクターブ低く歌唱していたことがあげられる．

図 2 はグレースケール画像を 2 値化してラベリングを適用した例である．この結果では，多くの歌唱者が有する同様な周波数推移に固有の色が割り当てられている．なお，この図ではマウスオーバーした箇所に該当する 1 人の歌唱者の周波数推移を白線の折れ線グラフとして同時に表示している．この図の楕円 (1) に着目すると，多くの人が同様な歌唱をしていることが紺のラベルで表示されているのに対して，高い周波数で赤のラベルが存在することがわかる．また楕円 (2) では，高い周波数での歌唱が緑のラベルで，低い周波数での歌唱が水色ラベルで表示されている．この可視化結果から，同様な歌い方の癖を持つ歌唱者のグループがいくるかある可能性が考えられる．

4. 一般的な楽曲音源からのデータ生成

前章で紹介した実行例では，多数の歌唱者によるデータの例として，最初から無伴奏歌唱音源として公開されているオープンデータを用いた．しかし，このようなオープンデータを幅広く入手できるとは限らない．伴奏とあわせて録音された一般的な歌唱音源から無伴奏データを生成することで，幅広い楽曲に対して可視化が可能になる．そこで，以下の 2 つの手法を用いて伴奏付き歌唱音源から無伴奏歌唱音源を抽出した．この処理によって生成された無伴奏歌唱音源の F0 を推定することで，幅広い楽曲への適用が可能となる．

以下，減算処理または Spleeter[32] による無伴奏歌唱音源の抽出，歌唱音源からの F0 推定，の順に論じる．

4.1 減算処理による抽出

4.1.1 時刻及びキーオフセット推定

まず，伴奏付き歌唱音源と伴奏音源を Constant–Q 変換によりスペクトログラムに変換する．この 2 音源の時間と周波数の二次元配列の相互相関の最大値を求めることにより，音源の始まる時刻やキーのずれを検出する．ここで，相互相関を求める範囲には歌唱が含まれていないことが望ましいため，多くの楽曲において前奏となる「音源開始から 10 秒程度」を用いる．ただし，「サビ始まり」の歌唱から始まる楽曲においては，音源開始以外の箇所から 10 秒程度を用いる．

4.1.2 音量オフセット推定

続いて，伴奏差分によって音量オフセットを推定する．時刻のずれを修正した伴奏付き歌唱音源の振幅スペクトルと，伴奏音源から STFT によって変換した振幅スペクトルを，それぞれ $S1$，$S2$ とする．この二つの音源は同じ音量であるとは限らないため，無伴奏歌唱音源の振幅スペクトル $S3$ を

$$S3 = S1 - \Gamma * S2 \tag{5}$$

で求める．ここで Γ の値を各音源について推定する必要がある．Γ の値が大きすぎれば，伴奏を引いた結果として負値になる箇所が増える．この負になった箇所の割合が閾値を超えない最大値を Γ とする．

4.2 Spleeter による抽出

4.2.1 時刻及びキーオフセット推定

Spleeter を用いる場合にも，4.1.1 節にて示した手法により音源の始まる時刻のずれを検出し，音源の開始時刻を揃える．ここで，キーが異なる音源については別のファイルに書き出す．

4.2.2 Spleeter

Spleeter[32] は，Image-to-Image の分野で提案された U-Net 構造 [31] を音源分離に適用した手法である．これを用いることでも無伴奏歌唱音源を抽出できる．

4.3 F0 推定

減算処理または Spleeter を用いて抽出した歌唱音源に対して，F0 を推定する．上述の 2 つの手法で得られる無伴奏歌唱音源には抑制しきれない雑音が残ることがあるため，耐雑音性に優れた F0 推定手法を適用することが望ましい．本報告では PYIN[33] を用いた．

4.4 正解音高推移の表示を追加実装した可視化結果

前節までに示した手法で F0 推定した結果を可視化した．可視化の実行環境は 4 章と同一である．歌唱データには “夜明けと蛍” の 67 人の歌唱を用いた．この曲は VOCALOID を用いて原曲が公開された楽曲である．歌唱の音高推移を可視化するにあたり，原曲の歌声合成に用いられた楽譜相当の音高推移データを正解音高推移として同時に表示する機能を追加実装した．これにより，歌唱者群がどの程度正確な歌唱を実現できているか，あるいは逆に正解データの音高と差のある音高を意図的に歌唱する唱法がどの瞬間に用いられているか，といった点を観察できるようになる．

図 3 に音高推移の分布を可視化した例を示す．楽曲の途中から，上下 2 か所にわたって同様な音高推移が見られることがわかる．この可視化結果に対して, 図 4 に示す正解音高推移を重ねた例を図 5 に示す．楽曲のサビにあたる箇所で，原曲の音高は上に見られるクラスタの方へ推移しており，音域が大きく上昇することがわかる．そのため可視化結果からは半分ほどの歌唱者がサビで 1 オクターブ低く歌っていることが読み取れ，実際に 30 人がサビを 1 オクターブ下げて歌っていた．またサビ部分にあたる 1 オクターブ低くした音高推移は，A メロにあたる箇所の音高推移とほぼ同じ音域に属していることから，この楽曲はサビを 1 オクターブ下げてもさほど違和感のない曲であること

も推察される．

図 6 は図 3 を 2 値化しラベリングした例である．図 3 と比較して，ラベリングによって新たに特徴的な F0 推移を見つけることは難しい．図 2 とは異なりデータの歌唱者数が少ないことから，ラベリングによる可視化手法の調整が必要であると考えられる．

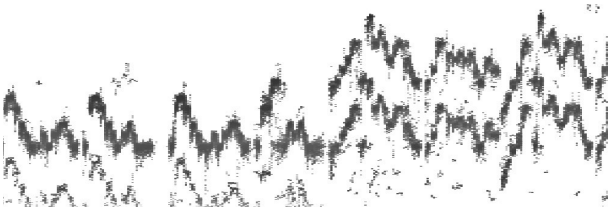

図 3 音高推移分布をグレースケール画像として表示した例.

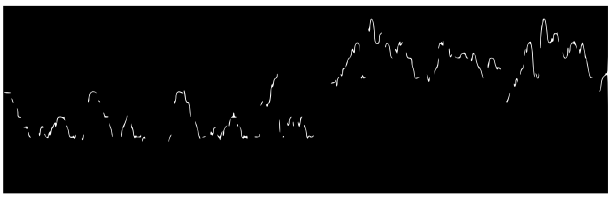

図 4 原曲の音高推移

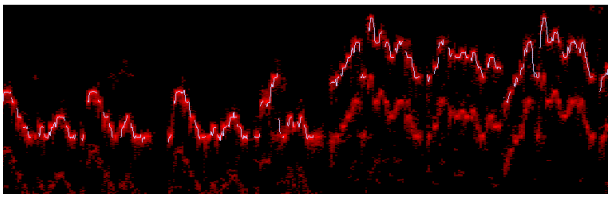

図 5 図 3 に対し原曲を正解音高として白線で表示した例.

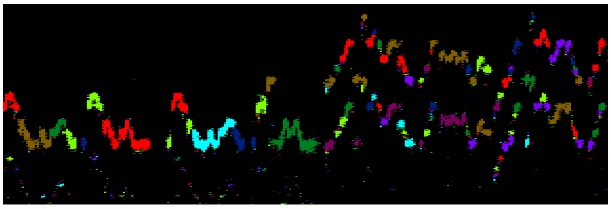

図 6 グレースケール画像を 2 値化してラベリングした例.

5. 今後の展望

本研究に対する今後の展望として，以下のようなことを検討している．

- **付随情報の同時表示**
  歌唱者の属性と音高推移との関係性を調査する．例えば歌唱者の年齢や歌唱経験年数，あるいは web 上で公開されている歌唱であれば評価や再生数，といった属性を各歌唱データに付与させることが可能である．これらの属性と音高推移との関係が明らかになれば興味深い．あるいは，歌唱技法と音高推移の関係を可視化することも考えられる．音高推移から評価と数値化がある程度可能な歌唱技法の例として，ビブラート [9] やオーバーシュートがあげられる．つまり，周波数以外にも変数を有する時系列データとして歌唱群を可視化することが今後の課題の一つである．
- **可視化手法の改善**
  図 1,3 からもわかるように，現在実装されている密度関数のヒートマップを適用した可視化では，個々の歌唱の音高推移を鮮明に視認することが難しい．この解決方法として，密度関数のヒートマップをラベリング処理だけのために用いて，個々の歌唱の音高遷移を折れ線グラフや散布図で表示する方法が考えられる．一方で 2 章で議論した通り，折れ線グラフや散布図には Visual Cluttering という問題が避けられない．これを解決する手段として，ネットワークや多次元データの可視化において既に多用されている Bundling(束化)[30] という手法を用いることができる．さらに別の問題として，現状の実装では，多数の歌唱者で頻出する音高推移パターンを発見することには向いているが，例外的な音高推移を発見するのが難しいという点もあげられる．この点についても，ラベリング処理から外れた音高推移を折れ線グラフや散布図で表示するという形での解決可能性がある．
- **可視化結果からの集合知の抽出**
  多数の楽曲の歌唱データに対して本手法を適用することで，歌唱群の音高推移に対する多数の可視化結果を集めることができる．VOCALOID 等の歌声合成技術を用いることで，同一楽曲に対して異なる制作者が歌唱作品を制作して動画共有サービス等で公開する文化があり，それらを可視化して比較することは興味深い．そのようにして得られた複数の可視化結果に対して機械学習を適用することで，「このような音高推移が見られる楽曲に対して，このような音高推移での歌唱がよくみられる」といった一般性の高い歌唱分析が可能になるであろうと期待する．
- **制作者支援ツール**
  歌唱者が自分の歌唱技術や表現力を向上させるための支援ツールとして，あるいは歌声合成技術を用いる制作者が他者の制作技術を参考にするための支援ツールとして，本手法がどのように貢献できるかを実証したい．

6. まとめ

我々は，同一楽曲に対する多数の歌唱データに対して推定した音高の推移を時系列データとみなし，画像処理的なアプローチによって可視化する手法を提案している．本報告ではその可視化手法と実行例を示し，さらに一般的な楽

曲音源からのデータ生成手法について述べた．今後は，前章で示した展望に沿って，さらに研究開発を進めたい．

参考文献

[1] M. Last, A. Kandel, H. Bunke, Data Mining in Time Series Databases, World Science Publishing, ISBN-981-238-290-9, 2004.
[2] T. W. Liao, Clustering of Time Series Data — A Survey, Pattern Recognition, 38, pp.1857-–1874, 2005.
[3] 伊藤 貴之, 中野 倫靖, 深山 覚, 濱崎 雅弘, 後藤 真孝, 同一楽曲に対する多数の歌唱の基本周波数推定値分布の可視化, 情報処理学会研究報告音楽情報科学 (MUS), 2019-MUS-123, 47, pp.1–6, 2019.
[4] J. Wilkins, P. Seetharaman, A. Wahl, B. Pardo, Vocalset: A Singing Voice Dataset, Proc. ISMIR 2018, 2018.
[5] K. Tsuzuki, T. Nakano, M. Goto, T. Yamada, S. Makino, Unisoner: An Interactive Interface for Derivative Chorus Creation from Various Singing Voices on the Web, Proc. Joint ICMC SMC 2014 Conference, 2014.
[6] 都築 圭太, 中野 倫靖, 後藤 真孝, 山田 武志, 牧野 昭二, Unisoner：様々な歌手が同一楽曲を歌った Web 上の多様な歌声を活用する合唱制作支援インタフェース, 情報処理学会論文誌, 56(12), pp.2370–2383, 2015.
[7] S. Wager, G. Tzanetakis, S. Sullivan, C. Wang, J. Shimmin, M. Kim1, P. Cook, INTONATION: A Dataset of Quality Vocal Performances Refined by Spectral Clustering on Pitch Congruence, Proc. ICASSP 2019, pp.476–480, 2019.
[8] 中野 倫靖, 後藤真孝, 平賀譲, 楽譜情報を用いない歌唱力自動評価手法, 情報処理学会論文誌, 48 (1), pp.227–236, 2007.
[9] J. Driedger, S. Balke, S. Ewert, M. Muller, Template-Based Vibrato Analysis in Complex Music Signals, Proc. ISMIR 2016, 2016.
[10] 後藤 真孝, 齋藤 毅, 中野 倫靖, 藤原 弘将, 歌声情報処理の最近の研究, 日本音響学会誌, 64 (10), pp.616–623, 2008.
[11] D. Hoppe, M. Sadakata, P. Desain, Development of Real-time Visual Feedback Assistance in Singing Training: A Review, Journal of computer assisted learning, 22(12), pp.308—316, 2006.
[12] T. Nakano, M. Goto, Y. Hiraga, MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data, Proc. the IEEE International Symposium on Multimedia (ISM 2007) Workshops, 2007.
[13] M. Shiraishi, K. Ogasawara, T. Kitahara, HAMOKARA: A System for Practice of Backing Vocals for Karaoke, Proc. SMC 2018, pp.511–518, 2018.
[14] F. Moschos, A. Georgaki, G. Kouroupetroglou, FONASKEIN: An Interactive Software Application for the Practice of the Singing Voice, Proc. SMC 2016, pp.326–331, 2016.
[15] O. Mayor, J. Bonada, A. Loscos, Performance Analysis and Scoring of the Singing Voice, Proc. AES 35th International Conference, 2009.
[16] T. Nakano, M. Goto, VocaRefiner: An Interactive Singing Recording System with Integration of Multiple Singing Recordings, Proc. SMC 2013, pp.115–122, 2013.
[17] A. Klapuri, A Method for Visualizing the Pitch Content of Polyphonic Music Signals, Proc. ISMIR 2009, 2009.
[18] L. Jure, E. Lopez, M. Rocamora, P. Cancela, H. Sponton, I. Irigaray, Pitch Content Visualization Tools for Music Performance Analysis, Proc. ISMIR 2012, 2012.
[19] E. Gomez, J. Bonada, Tonality Visualization of Polyphonic Audio, Proc. ICMC 2005, 2005.
[20] K. W. E. Lin, H. Anderson, N. Agus, C. So, S. Lui, Visualising Singing Style Under Common Musical Events Using Pitch-Dynamics Trajectories and Modified TRACLUS Clustering, Proc. ICMLA 2014, 2014.
[21] T. Kako, Y. Ohishi, H. Kameoka, K. Kashino, K. Takeda, Automatic Identification for Singing Style Based on Sung Melodic Contour Characterized in Phase Plane, Proc. ISMIR 2009, 2009.
[22] Y. Uchida, T. Itoh, A Visualization and Level-of-Detail Control Techniquefor Large Scale Time Series Data, Proc. IV09, pp.80–85, 2009.
[23] C. Perin, F. Vernier, J.-D. Fekete, Interactive Horizon Graphs: Improving the Compact Visualization of Multiple Time Series, Proc. ACM CHI 2013, pp.3217–3226, 2013.
[24] Y. Wang, F. Han, L. Zhu, O. Deussen, B. Chen, Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series, IEEE Trans. on Visualization and Computer Graphics, 24(2), pp.1141–1154, 2018.
[25] M. Imoto, T. Itoh, A 3D Visualization Technique for Large Scale Time-Varying Data, Proc. IV10, pp.17–22, 2010.
[26] G. Oliveira, J. Comba, R. Torchelsen, M. Padilha, C. Silva, Visualizing Running Races through the Multivariate Time-Series of Multiple Runners, Conference on Graphics, Patterns and Images, 2013.
[27] R. Mazza, Introduction to Information Visualization, Springer, ISBN:978-1-84800-218-0, 2009.
[28] J. J. van Wijk, E. W. van Selow, Cluster and Calendar based Visualization of Time Series Data, Proc. InfoVis'99, 1999.
[29] H. Kawahara, I. Masuda-Katsuse, A. de Cheveigne, Restructuring Speech Representations Using a Pitch Adaptive Time-frequency Smoothing and an Instantaneous Frequency Based on F0 Extraction: Possible Role of a Repetitive Structure in Sounds, Speech Communication, 27, pp.187-–207, 1999.
[30] D. Holten, Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data, IEEE Trans. on Visualization and Computer Graphics, 12(5), pp.741–748, 2006.
[31] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, CoRR, Vol. abs/1505.04597 , 2015.
[32] R. Hennequin, A. Khlif, F. Voituret, M. Moussallam, Spleeter: A Fast And State- of-the Art Music Source Separation Tool With Pre-trained Models, Late-Breaking/Demo ISMIR 2019 by Deezer Research, 2019.
[33] M. Mauch, S. Dixon, PYIN: A fundamental fre-quency estimator using probabilistic threshold distribu-tions, Proc. ICASSP 2014, pp.659-–663, 2014.