

同一楽曲に対する多数の歌唱の基本周波数分布の マルチスケール可視化

伊藤 貴之^{1,a)} 中野 倫靖^{2,b)} 深山 覚^{2,c)} 濱崎 雅弘^{2,d)} 後藤 真孝^{2,e)}

概要: 2次創作の流行や歌声情報処理の発達により、多数の歌唱者が同一の楽曲を歌った音源を耳にする機会が増えた。このような歌唱群の傾向を理解する一手段として我々は、同一楽曲に対する多数の歌唱群の音響データから抽出された音高（基本周波数：F0）の推移を可視化する手法を開発している。本報告ではその続報として、楽曲全体・声域全体にわたる概略的な分布を示す画面と、その中の局所をズームアップした画面の組み合わせによって、歌唱群のF0推移の全体的傾向および局所的傾向を表現するマルチスケールな可視化手法を提案する。本報告では1000人の歌唱者による同一楽曲の歌唱と、その各歌唱に対する5段階評価結果を可視化した事例を示し、本手法の有用性を検証する。

1. はじめに

近年の2次創作の流行や歌声情報処理の発達にもない、同一楽曲に対する多数の歌唱を耳にする機会が増えている。このような多数の歌唱データの分析・比較は学術的にも興味深い。例えば、多数の歌唱者による同一楽曲の歌唱を比較することで、どのような癖や個性を有する歌唱者が多い傾向にあるか、楽譜上の音高（基本周波数：以下F0と称する）からどのようにずれた表現をする歌唱者が多い傾向にあるかを分析できる[1]。また、高評価な歌唱者と低評価な歌唱者の歌唱を比較することで、高評価な歌唱にはどのような表現がよく見られるか、といった議論も可能になる。あるいは歌声合成による歌唱データの場合には、どのような表現手段を用いているデータが多いか、あるいはどのような表現手段を用いている歌唱が注目される傾向にあるか、といったパターンを分析できる。

F0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[2,3]、それらを適用することで歌唱の違いや逸脱を検出することは可能である。一方で歌唱の分析には主観的な解釈を要する場面もあり得る。例えば同

一楽曲の特定の瞬間のF0に個人差が見られた際に、意図的な歌唱技法としてF0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F0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な場合がある。そこで我々は、分析者が主観的・定性的に歌唱データを観察するための可視化手法の開発に着手している[4,5]。可視化は大規模なデータの理解の目的で多用されており、本研究の問題設定である「多数の歌唱データの理解」にも合致している。また鑑賞するのに一定の時間を要する音楽のようなデータをより短時間で理解できるという意味において、可視化は歌声のF0を分析する際に有用な手段である。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。

本報告では「Overview+Detail」[6]という可視化デザインに沿って、以下の2画面

Overview: 楽曲全体・歌唱声域全体にわたる概略的なF0分布をグレースケール画像で表現する画面。

Detail: 数ヘルツ単位の細かいF0分布、数秒単位の局所的なF0分布を折れ線の集合で表現する画面。
に沿ったマルチスケールな可視化手法を提案する。

図1に本研究が想定する提案手法の操作手順を示す。ユーザはまずOverview画面から例えば、あえて異なる音階で発声している歌唱者がいるフレーズや、幅広い周波数帯域に分散して歌唱が分布するフレーズが、楽曲全体の中でどのように存在しているかを観察する。その上で画面中の興味ある矩形領域を指定すると、Detail画面にてその中に分布する歌唱群の微細なF0分布や、歌唱評価との関係

¹ お茶の水女子大学
東京都文京区大塚 2-1-1

² 産業技術総合研究所
茨城県つくば市梅園 1-1-1

a) itot@is.ocha.ac.jp

b) t.nakano@aist.go.jp

c) s.fukayama@aist.go.jp

d) masahiro.hamasaki@aist.go.jp

e) m.goto@aist.go.jp

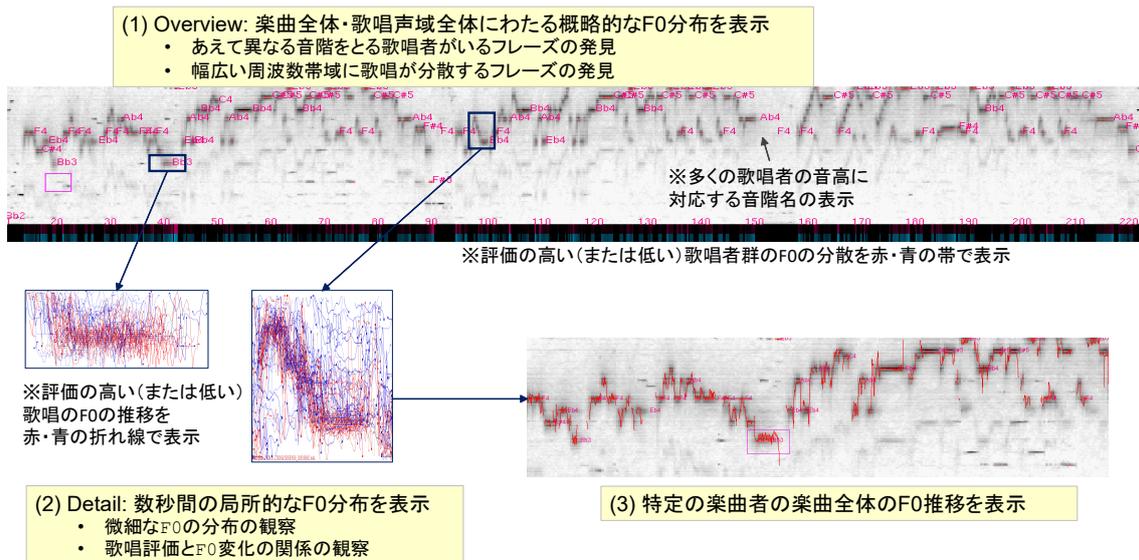


図 1 提案手法のスナップショットおよび想定する操作手順。提案手法ではまず (1) に示す Overview 画面にて、楽曲全体・歌唱声域全体にわたる F0 分布をヒートマップで表示する。ここからユーザーが特定の矩形領域を指定すると、(2) に示す Detail 画面にて、矩形領域に該当する歌唱者の F0 遷移を折れ線で表示する。赤い折れ線は高評価を、青い折れ線は低評価を示している。ここから特定の折れ線を指定すると、Overview 画面にて当該歌唱者の F0 遷移を楽曲全体にわたって折れ線で表示する。

を観察できる。さらに、ここで特定の歌唱を指定すると、Overview 画面にてその歌唱者の楽曲全体の F0 推移を表示することで、その歌唱者の楽曲全体にわたる傾向を観察できる。

本手法では各歌唱について一定時刻ごとに推定した F0 の集合を時系列データとする。Overview では、横軸を時刻・縦軸を周波数の対数値とする格子から生成した 2 次元ヒストグラムをグレースケール画像として表示する。この Overview から矩形領域を指定すると、Detail ではその矩形領域内部に該当する F0 推移を折れ線の集合で表現する。

本報告では 1000 人の歌唱者による同一楽曲の F0 分布を可視化した例を示し、その有効性を検証する。

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 [7] 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、同一楽曲に対する多数の歌唱を活用した 2 次創作によって、合唱や輪唱などの新しい作品を制作することができる。歌唱作品の制作環境の一例として都築ら [8,9] は、そうした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 F0 推移の可視化

演奏データにおける F0 推移の分析や観察に可視化を用いた研究事例には、Nakano ら [10] の MiruSinger, Shiraishi ら [11] の HAMOKARA, Moschos ら [12] の FONASKEIN, Mayor ら [13] の歌唱採点手法などがあげられる。複数の歌唱を対象とした可視化の事例として、Nakano ら [14] 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 F0 の可視化も採用している。

他には、演奏情報の中から F0 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 [15,16] や、周波数情報から推察される調性の可視化 [17] などの事例がある。また、歌唱の F0 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、F0 とダイナミクスを 2 軸とした可視化 [18] や、F0 と F0 差分を 2 軸とした可視化 [19] が報告されている。Wilkins ら [7] 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。また Weiß ら [20] は、録音された合唱のイントネーション品質を測定するために顕著性の高い周波数群を可視化している。

これらの研究に対して我々の研究 [4,5] は、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ために Overview+Detail 型の可視化手法を導入する、という点で独自性を有する。

2.3 時系列データの可視化

歌唱の F0 の推移は時系列データとして扱うことが可能

である。 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と、以下の3種類の可視化手法の適用が可能である。

- (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 [21–23] や散布図 [23]。
- (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。時刻と実数値を2軸とした空間でのヒートマップという意味ではスペクトログラムに近い視覚表現である。
- (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・色相 [24, 25] あるいは線分の太さ [26] で表現したもの。

近年では時系列データのための各種の可視化手法に関する主観比較評価 [27] や類似度認知実験 [28] も発表されている。我々も本研究の Overview 画面に関する先行研究 [4] にて、上記の3種類の可視化手法を試行した結果を比較し、その問題点を指摘している。(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。一方で (3) に示したヒートマップでは、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 [29] ことが実用上の問題となる。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」を Overview 画面に採用している。一方で Detail 画面においては表示対象となる折れ線の本数が限定的であり、Visual Clutter の影響が小さいことから、本研究では (1) に示す折れ線グラフを採用している。

時系列データの分析では大局的な時間変化と局所的な時間変化の両方に注視すべき事例が多い。そのため、大局的な表現と局所的な表現を組合せたマルチスケールな可視化手法の適用が有用な場面が多い。このような考えにもとづく可視化デザインは既に多くの研究に採用されており、以下のように大別される。

- (1) 時系列データを構成する同一の変数値・属性値に対して同一の視覚表現を適用したマルチスケール可視化 [30, 31]。
- (2) 時系列データを構成する同一の変数値・属性値に対して視覚表現を切り替えて適用したマルチスケール可視化。
- (3) 時系列データを構成する別々の変数値・属性値に対して視覚表現を切り替えて適用したマルチスケール可視化 [32, 33]。

これらの手法の中で本手法は、上記の (2) の考えに沿って、「F0」という同一の変数値に対して、大局的な表現にはヒートマップ、局所的な表現には折れ線グラフ、という異なる表現を適用して歌唱分布を表現した点に特徴がある。

3. F0 分布の可視化

3.1 前提となるデータ：F0 と歌唱評価結果

本章では歌唱者集合 S を構成する各歌唱者の F0 推移およびその評価を以下のように表記する。

$$S = \{s_1, s_2, \dots, s_n\}$$

$$s_i = \{p_{i1}, p_{i2}, \dots, p_{im}, e_i\} \quad (1)$$

ここで s_i は i 番目の歌唱者による歌唱の F0 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 の対数、 e_i は i 番目の歌唱者への評価値、 m は F0 推定の対象区間における標準化された時刻の総数（各 F0 系列の F0 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 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。この前提が成立しないデータが与えられた際には、DP マッチング等による歌唱間の F0 系列のアライメントが前処理として必要である。

3.2 F0 分布の性質と本研究が想定するタスク

本研究が扱う F0 分布の性質を以下の通り、時間分布と周波数分布にわけて考察する。

時間分布: 歌唱の F0 は楽譜に沿って分布する。よって楽曲全体にわたる F0 分布を可視化することで、楽曲の進行と F0 分布の関係を観察できる。また音符 1 個単位、あるいはもっと細かい数ミリ秒単位の F0 分布を可視化することで、特定の音階と F0 分布の関係や、ビブラートやオーバーシュートといった歌唱技法と F0 分布の関係を観察できる。つまり F0 分布の可視化は時間軸に対してマルチスケールな問題であると考えられる。

周波数分布: 歌唱の周波数は音階に沿って分布する。よって歌唱音域全体にわたる周波数分布を可視化することで、あえて異なる音階で発声している歌唱者がいるフレーズの発見や、幅広い周波数帯域に分散して歌唱が分布するフレーズの発見が可能になる。一方で、各瞬間における旋律本来の音階周辺の数ヘルツ単位の周波数帯域に着目することで、歌唱者ごとの細かい F0 の差を観察できる。つまり F0 分布の可視化は周波数軸に対してもマルチスケールな問題であると考えられる。

以上の考察にもとづき本報告では、「Overview」「Detail」の2画面から構成される可視化により以下のタスクを支援する手法を提案する。

Overview: あえて異なる音階で発声している歌唱者がいるフレーズや、幅広い周波数帯域に分散して歌唱が分布するフレーズが、楽曲全体の中でどのように存在しているかを観察する。そして注目に値するフレーズの

F0 分布を対話操作によって指定し、Detail にて表示させる。

Detail: 楽曲中の数秒単位の短時間を対象として、歌唱者間の数ヘルツ単位の微小な F0 の差異や、その分布と歌唱評価との関係を観察する。そして注目に値する特定の歌唱者を対話操作によって指定し、その歌唱者の楽曲全体にわたる F0 推移を Overview にて表示させる。

3.3 Overview : ヒートマップ上でのインタラクション

3.3.1 ヒートマップの生成

Overview では F0 推移の密度分布をヒートマップで表現する [4]。このヒートマップでは、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_{start} , t_{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 $t_{start} = t_1$, $t_{end} = t_{N+1}$, $p_{min} = f_1$, $p_{max} = f_{M+1}$ とする。

続いて、 $F0p_{ij}$ 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属するとして、変数 r_{uv} に 1 を加算する。

以上の処理によって構成される 2 次元ヒストグラムを横 N 画素、縦 M 画素のヒートマップ画像として扱う。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。ただし α および γ はユーザが調節可能な変数とする。図 2 に示す通り、 α が大きいほうが r_{uv} が大きい部位での明度が低く (= 黒が濃く) なり、 γ が大きいほうが r_{uv} が大きい部位と小さい部位のコントラストが大きくなる。

さらに Overview では、ヒートマップを構成する画素の各列のうち、音階を構成する各音の周波数に対応する画素を対象として、 r_{uv} の値が一定以上かつ最大である画素に別の色を割り当てる。これにより、最も多くの歌唱者がとっている音階をヒートマップ上で明示する。

3.3.2 ヒートマップ上での矩形領域操作

Overview には時刻の範囲 $[t_1, t_2]$ と F0 の範囲 $[p_L, p_H]$ を示す矩形領域 (図 3 参照) を重ねて表示する。この矩形領域はマウスのクリック操作により「可動モード」と「停止

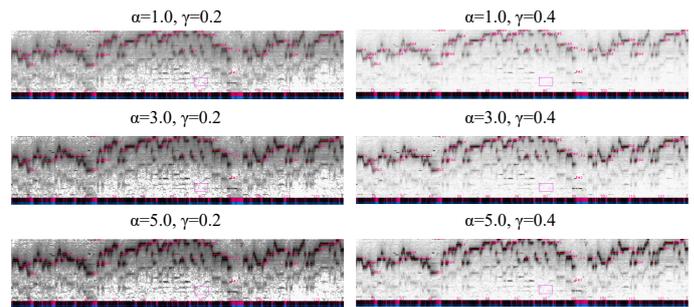


図 2 α および γ の調節に対する可視化結果の変化。

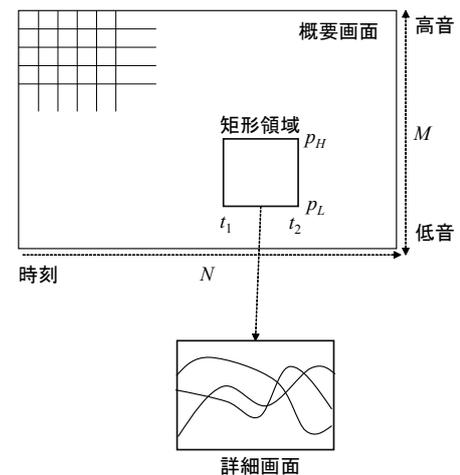


図 3 Overview における座標値の定義。

モード」が切り替えられる。可動モードではマウスのポインタ移動操作にともなって Overview を移動し、停止モードでは後述する Detail において矩形領域内部に対応する F0 推移を折れ線で表示する。

3.3.3 F0 分散の概略表示

さらに Overview では、各瞬間の F0 の分散もヒートマップで概略的に表示する。現時点での我々の実装では

- 評価の高い歌唱群の F0 の分散
- 評価の低い歌唱群の F0 の分散

を 2 色のヒートマップで表示する。これにより、「高評価な歌唱者があえて正確な F0 をとっていないフレーズ」「低評価な歌唱者が F0 を正しくとれない傾向があるフレーズ」などを発見しやすくなり、矩形領域を操作する際の有力なヒントにもなる。なお我々の実装では、歌唱評価は 5 段階のリッカート尺度で与えられているとし、評価 4 または 5 の歌唱群の F0 の分散を赤いヒートマップで、評価 1 または 2 の歌唱群の F0 の分散を青いヒートマップで表示している。

3.4 Detail : 本数と不透明度を制御した折れ線表示

Detail では Overview の矩形領域内部に対応する F0 推移を折れ線の集合で表現する。ここで多数の歌唱者の F0 推移の中から重要な知見を得るために、本手法では折れ線

の描画に際して以下の機能を搭載している。

- 折れ線の色と不透明度を歌唱者の評価から指定する。
- 発声の開始時と終了時を強調表示することで、各歌唱のタイミングの適切さを可視化する。
- 後述するサンプリングおよびクラスタリングの各手法により、描画する折れ線の本数を制御する。

発声の開始時と終了時を強調表示するために本手法では、 $p_{i(j-1)} = 0$ または $p_{i(j+1)} = 0$ を満たす p_{ij} の位置に一定の大きさの点を描画する。また本手法では、特定の折れ線をクリック操作で指定すると、その折れ線に対応する歌唱の F0 を、Overview にも折れ線で表示する。

描画する折れ線の本数を制御する 1 つ目の手法として、本手法ではサンプリング機能を搭載する。この機能では各々の折れ線に対して一様乱数となる値 $z(0.0 \leq z \leq 1.0)$ を算出し、これが

$$\alpha_x z > Z_{thres} \quad (4)$$

を満たすならば当該折れ線を描画し、さもなければ当該折れ線の描画を割愛する。ここで Z_{thres} は折れ線の本数を制御するためにユーザが指定する閾値であり、 α_x は表 1 に記載された不透明度のいずれかである。この値の選択により、低評価な歌唱の折れ線のみ、あるいは高評価な歌唱の折れ線のみを多めに描画するように本数を調節することが可能である。なお我々の実装では、Overview の矩形領域を停止モードにした瞬間、あるいは Z_{thres} の値を変更した瞬間にサンプリングを実行する。図 4 に Z_{thres} の値と折れ線の本数の関係を示す。

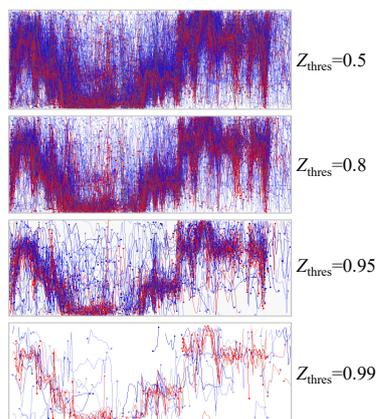


図 4 Z_{thres} の調節に対する折れ線の表示本数の変化。

描画する折れ線の本数を制御する 2 つ目の手法として、本手法ではクラスタリング機能を搭載する。ここでは描画対象となる折れ線群のうち i 番目と j 番目の折れ線について、時刻 t_1 から t_2 までの p_{ik} および p_{jk} の値を $(t_2 - t_1 + 1)$ 次元ベクトルとみなし、両者のマンハッタン距離 $dist_{ij}$ を以下の式で算出する。

$$dist_{ij} = \sum_{k=t_1}^{t_2} |p_{ik} - p_{jk}| \quad (5)$$

ただし $p_{ik} = 0$ または $p_{jk} = 0$ である場合には、十分大きな定数 $DIST$ *1 を設定し、 $|p_{ik} - p_{jk}| = DIST$ とする。このマンハッタン距離を全ての折れ線ペアに対して算出し、k-medoids 法を適用することで一定個数のクラスタを生成し、各クラスタの中心とされた折れ線を描画する。クラスタリングはサンプリングに比べて計算時間が大きいいため、画面上で所定のボタンを押したときのみ実行する。

4. 実行例

本手法を実装した結果を紹介する。著者らは本可視化手法を Java 1.12.0 および JOGL (Java binding for OpenGL) 2.3.2 で実装した。実行例には DAMP-balanced dataset *2 に収録された "Let It Go" の無伴奏歌唱の音響データから 1000 人の歌唱を無作為抽出して適用した。これらの歌唱に対して、1 人の評価者が各歌唱について 5 段階のリッカート尺度による主観評価を施した結果を付与した。評価者は "Let It Go" の各歌唱者に対して、1 コーラス目のサビ直前からそのサビ終了までを試聴した。さらに、Python 言語の楽曲解析ライブラリである librosa に搭載された F0 推定アルゴリズム pYIN を用いて、音響データから F0 を算出した。そして、相互相関が最大となるようなアライメントを適用することで、歌唱間のタイミングを補正した。可視化結果の画素数は $N = 1000, M = 480$ 、F0 の単位は Hz とし、対象となる周波数を 110Hz~1760Hz (音階で A1 から A5) の 4 オクターブとした。

楽曲全体を表す Overview と、矩形領域の操作にともなう Detail の表示例を図 5 に示す。この図の上半分が Overview であり、黒に近い色ほど多くの歌唱者が該当する音域で発声していることを示す。Overview の下端には、高評価および低評価である歌唱の F0 の分散を、それぞれ赤と青のヒートマップで表示している。ユーザはこのヒートマップを参照しながら、矩形領域の移動先を選ぶことができる。図 5(1)~(6) は Overview 中の 6 ヶ所に矩形領域を移動して表示した Detail の例である。

図 5 に示す実行結果では、Detail での歌唱者の評価と折れ線の色・不透明度の対応を表 1 のとおり設定している。この例では低評価ほど原色に近い青、高評価ほど原色に近い赤が割り当てられる。また不透明度に α_1 を選択すれば評価が「2」である歌唱を強調し、 α_2 を選択すれば評価が「5」である歌唱を強調し、 α_3 を選択すれば評価「2」「5」の両者を強調する。今回採用したデータでは評価が「1」である歌唱は分析対象に値しないものが多いと判断し、強調

*1 $DIST$ の値が大きいほど発声の開始・終了のタイミングが合っている歌唱間の距離が小さくなり、 $DIST$ の値が小さいほど両者ともに発声しているときの F0 が近い歌唱間の距離が小さくなる。

*2 <https://ccrma.stanford.edu/damp/>

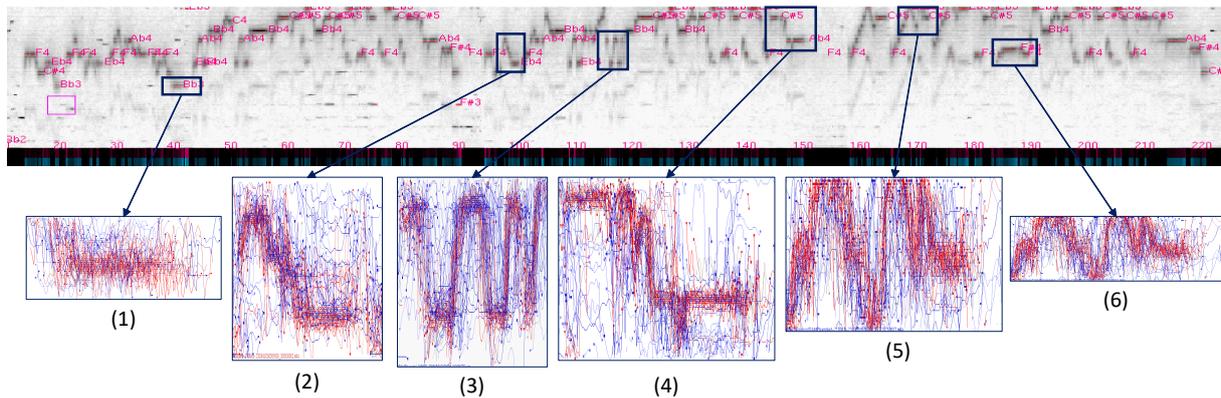


図 5 Overview 表示の例と、そこから 6 箇所を矩形表示を対象として Detail 表示を適用した例。Detail 表示では 5 段階評価で「5」となった歌唱を濃い赤で、「2」となった歌唱を濃い青で表示している。

表示しないことにした。

図 5(1)~(6) は α_3 を選択した表示結果であり、評価「2」「5」の両者がそれぞれ青、赤で強調表示されている。なお、図 5(1)~(6) の各々に対して別々のサンプリング結果が適用されているため、評価「2」の歌唱者が多く表示される場合と評価「5」の歌唱者が多く表示される場合がある点に注意されたい。

表 1 歌唱者の評価と折れ線の色・不透明度の対応の例

評価	R	G	B	α_1	α_2	α_3
1	0.0	0.0	1.0	0.5	0.5	0.5
2	0.1	0.1	0.8	1.0	0.3	1.0
3	0.4	0.4	0.4	0.5	0.5	0.5
4	0.8	0.1	0.1	0.4	0.8	0.8
5	1.0	0.0	0.0	0.3	1.0	1.0

図 5(1)(2) はいずれも、音階が下降した後にロングトーンがある。図 5(1) は赤い折れ線が比較的太い帯を形成しており、高評価な歌唱であっても多くの人が少しずつ異なる周波数で発声していることがわかる。それに対して図 5(2) は、少数の高評価の歌唱者が半音単位で異なる F0 で発声していることがわかる。また低評価の歌唱者がそれ以上に大きく F0 を外しているのがわかる。これらの可視化結果から、音階が下降するフレーズといっても異なる F0 分布が見られることがわかる。ただし図 5(1) に関して言えば、F0 を正確に制御できない歌唱者が多いのか、ビブラートなどの唱法によって F0 が上下する歌唱者が多いのか、といった点まではこの可視化結果だけでは判断できない。

図 5(3)(4)(5) はいずれも、音階が大きく上下するフレーズを示したものである。高評価な歌唱者は正確に F0 をとれており、低評価な歌唱者は正確に F0 をとれていない、という差が顕著にあらわれている。また図 5(5)(6) では、図 5(1) と同様に、フレーズの末尾のロングトーンでは高評価な歌唱者であっても F0 に幅があることがわかる。これも図 5(1) と同じく、F0 を正確に制御できない歌唱者が多い

のか、ビブラートなどの唱法によって F0 が上下する歌唱者が多いのか、といった点まではこの可視化結果だけでは判断できない。

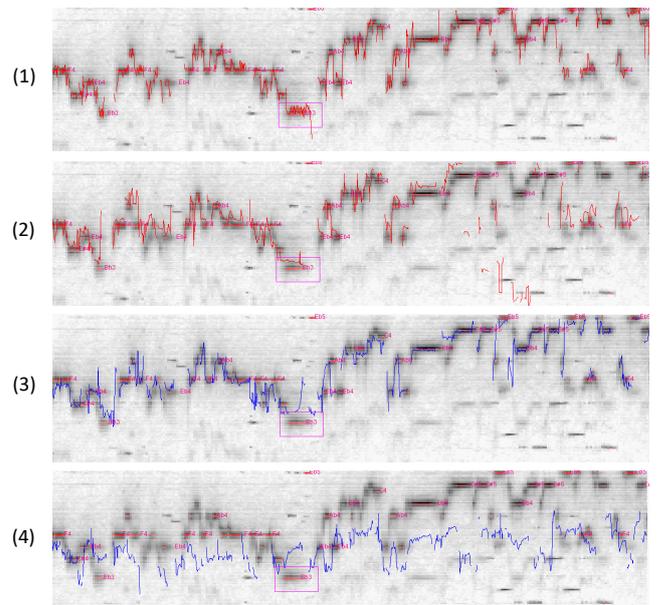


図 6 特定の歌唱者の F0 推移の表示例。(1) 多数の歌唱者と同様な F0 推移をとっており評価「5」となった歌唱者。(2) 多数の歌唱者と異なる F0 推移をとりながらも評価「5」となった歌唱者。(3) 多数の歌唱者と同様な F0 推移をとりながらも評価「2」となった歌唱者。(4) 多数の歌唱者と異なる F0 推移をとっており評価「2」となった歌唱者。

図 5(1) に示す Detail から特定の折れ線をクリックして、対応する歌唱を Overview にて折れ線表示した例を図 6 に示す。図 6(1) は図 5(1) 中で赤い折れ線が集中している部分から抽出した歌唱であり、楽曲全体にわたって正確な F0 で歌唱していることが見受けられる。この歌唱者に限らず、図 5(1) 中で赤い折れ線で表示されている高評価な歌唱の多くに、同様な F0 推移が見られた。それに対して少

数ながら、図 6(2) に示すように、多数の歌唱者とは異なる F0 推移を有しながらも高評価を得ている歌唱もある。一方で低評価な歌唱の中には、多数の歌唱者の F0 推移に近い図 6(3) のような F0 推移を有する歌唱と、多数の歌唱者の F0 推移から大きく離れた図 6(4) のような F0 推移を有する歌唱とが同程度に見られた。

5. まとめ・今後の課題

本報告では同一楽曲に対する多数の歌唱の F0 分布を「Overview と Detail」という 2 つの画面で探索的に可視化する手法を提案し、その実行例を示した。今後の課題として以下を検討している。

歌唱データ処理の改善. pYIN によって算出した現状の F0 にはノイズと思われる値が随所に含まれている。これを削除することで、より滑らかに F0 推移を表現したい。また現状では各音素の発音の瞬間を認識していないため、連続的な発音の開始時と終了時のみを強調表示している。この点を拡張することで、各音素のタイミングも可視化したい。

実用シナリオに沿った可視化手法の拡張. 現状の可視化手法は「F0 分布の大局的および局所的な傾向を表現することだけを目指しており、特定の有用シナリオを想定したものではない。今後の機能拡張に際して、何らかの有用シナリオを想定し、それを満たす方向で機能を拡張したい。

歌唱以外への可視化手法の適用. 本手法は「時間軸と物理値軸の双方にマルチスケールな描画が必要」であり「欠損値（歌唱の場合には休符）が随所に含まれる」というデータに一般的に利用可能な可視化手法である。このようなデータは歌唱およびそれ以外の各種楽器の演奏情報に限定されない。音楽情報以外も含めて多様な時系列データの可視化への適用が可能であると考えられる。

参考文献

- [1] Wager, S., Tzanetakis, G., Sullivan, S., C. Wang, J. S., Kim, M. and Cook, P.: Intonation: A Dataset of Quality Vocal Performances Refined by Spectral Clustering on Pitch Congruence, *Proc. IEEE ICASSP 2019*, pp. 476–480 (2019).
- [2] Last, M., Kandel, A. and Bunke, H.: *Data Mining in Time Series Databases*, World Science Publishing (2004).
- [3] Liao, T. W.: Clustering of Time Series Data – A Survey, *Pattern Recognition*, Vol. 38, pp. 1857–1874 (2005).
- [4] 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, 情報処理学会音楽情報科学研究報告, pp. 2019–MUS–123 (2019).
- [5] 近藤芽衣, 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と正解歌唱の音高推移分布の可視化, 情報処理学会インタラクティブ 2021, pp. 3P–03 (2021).
- [6] Cockburn, A., Karlson, A. and Bederson, B. B.: A review of overview+detail, zooming, and focus+context in-

- terfaces, *ACM Computing Surveys*, Vol. 41, No. 2, pp. 1–31 (2009).
- [7] Wilkins, J., Seetharaman, P., Wahl, A. and Pardo, B.: VocalSet: A Singing Voice Dataset, *Proc. ISMIR 2018*, pp. 468–474 (2018).
- [8] Tsuzuki, K., Nakano, T., Goto, M., Yamada, T. and Makino, S.: Unisoner: An Interactive Interface for Derivative Chorus Creation from Various Singing Voices on the Web, *Proc. Joint ICMC SMC 2014 Conference*, pp. 790–797 (2014).
- [9] 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, Vol. 56, No. 12, pp. 2370–2383 (2015).
- [10] 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, Vol. 48, No. 1, pp. 227–236 (2007).
- [11] Shiraiishi, M., Ogasawara, K. and Kitahara, T.: HAMOKARA: A System for Practice of Backing Vocals for Karaoke, *Proc. SMC 2018*, pp. 511–518 (2018).
- [12] Moschos, F., Georgaki, A. and Kouroupetroglou, G.: FONASKEIN: An Interactive Software Application for the Practice of the Singing Voice, *Proc. SMC 2016*, pp. 326–331 (2016).
- [13] Mayor, O., Bonada, J. and Lascos, A.: Performance Analysis and Scoring of the Singing Voice, *Proc. AES 35th International Conference* (2009).
- [14] Nakano, T. and Goto, M.: VocaRefiner: An Interactive Singing Recording System with Integration of Multiple Singing Recordings, *Proc. SMC 2013*, pp. 115–122 (2013).
- [15] Klapuri, A.: A Method for Visualizing the Pitch Content of Polyphonic Music Signals, *Proc. ISMIR 2009*, pp. 615–620 (2009).
- [16] Jure, L., Lopez, E., Rocamora, M., Cancela, P., Sponton, H. and Irigaray, I.: Pitch Content Visualization Tools for Music Performance Analysis, *Proc. ISMIR 2012*, pp. 493–498 (2012).
- [17] Gomez, E. and Bonada, J.: Tonality Visualization of Polyphonic Audio, *Proc. ICMC 2005* (2005).
- [18] Lin, K. W. E., Anderson, H., Agus, N., So, C. and Lui, S.: Visualising Singing Style Under Common Musical Events Using Pitch-Dynamics Trajectories and Modified TRACCLUS Clustering, *Proc. ICMLA '14* (2014).
- [19] Kako, T., Ohishi, Y., Kameoka, H., Kashino, K. and Takeda, K.: Automatic Identification for Singing Style Based on Sung Melodic Contour Characterized in Phase Plane, *Proc. ISMIR 2009* (2009).
- [20] Weiß, C., Schlecht, S. J., Rosenzweig, S. and Müller, M.: Towards Measuring Intonation Quality of Choir Recordings: A Case Study on Bruckner's Locus Iste, *Proc. ISMIR 2019*, pp. 276–283 (2019).
- [21] Uchida, Y. and Itoh, T.: A Visualization and Level-of-Detail Control Technique for Large Scale Time Series Data, *Proc. IV09*, pp. 80–85 (2009).
- [22] Perin, C., Vernier, F. and Fekete, J.-D.: Interactive Horizon Graphs: Improving the Compact Visualization of Multiple Time Series, *Proc. ACM CHI 2013*, pp. 3217–3226 (2013).
- [23] Wang, Y., Han, F., Zhu, L., Deussen, O. and Chen, B.: Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series, *IEEE Trans. on Visualization and Computer Graphics*, Vol. 24, No. 2, pp. 1141–1154 (2018).
- [24] Imoto, M. and Itoh, T.: A 3D Visualization Technique

- for Large Scale Time-Varying Data, *Proc. IV10*, pp. 17–22 (2010).
- [25] Oliveira, G., Comba, J., Torchelsen, R., Padilha, M. and Silva, C.: Visualizing Running Races through the Multivariate Time-Series of Multiple Runners, *Proc. SIB-GRAPI 2018*, pp. 99–106 (2013).
- [26] Krstajic, M., Bertini, E. and Keim, D. A.: CloudLines: Compact Display of Event Episodes in Multiple Time-Series, *IEEE Trans. on Visualization and Computer Graphics*, Vol. 17, No. 12, pp. 2432–2439 (2011).
- [27] Javed, W., McDonnel, B. and Elmqvist, N.: Graphical Perception of Multiple Time Series, *IEEE Trans. on Visualization and Computer Graphics*, Vol. 16, No. 6, pp. 927–934 (2010).
- [28] Gogolou, A., Tsandilas, T., Palpanas, T. and Bezerianos, A.: Comparing Similarity Perception in Time Series Visualizations, *IEEE Trans. on Visualization and Computer Graphics*, Vol. 25, No. 1, pp. 523–533 (2019).
- [29] Mazza, R.: *Introduction to Information Visualization*, Springer (2009).
- [30] Cuenca, E., Sallaberry, A., Wang, F. Y. and Poncet, P.: MultiStream: A Multiresolution Streamgraph Approach to Explore Hierarchical Time Series, *IEEE Trans. on Visualization and Computer Graphics*, Vol. 24, No. 12, pp. 3160–3173 (2018).
- [31] Sjölander, K. and Beskow, J.: Wavesurfer - An Open Source Speech Tool, *Proc. International Conference on Spoken Language Processing (ICSLP 2000)* (2000).
- [32] Goto, M., Yoshii, K., Fujihara, H., Mauch, M. and Nakano, T.: SONGLE: A Web Service for Active Music Listening Improved by User Contributions, *Proc. ISMIR 2011*, pp. 311–316 (2011).
- [33] André, P., Wilson, M. L., Russell, A., Smith, D. A., Owens, A. and Schraefel, M. C.: Continuum: Designing Timelines for Hierarchies, Relationships and Scale, *Proc. UIST 2007*, pp. 101–110 (2007).