

SingDist Vis: 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

伊藤 貴之* 中野 倫靖 深山 覚 濱崎 雅弘 後藤 真孝†

概要. 人々が自身の歌った動画を動画共有サービスへ投稿する機会が増え、同一楽曲を異なる人々がそれぞれ歌った「同曲異唱コンテンツ」を楽しむことができるようになった。このような歌唱群の傾向を理解する一手段として、歌い方としての基本周波数 (F0) 推移の分布をインタラクティブに可視化できる歌声情報可視化インタフェース SingDist Vis を提案する。SingDist Vis は、曲全体を時間-周波数平面上で概略的に可視化する Overview 機能と、その局所部分をズームアップして歌唱力とともに可視化する Detail 機能の組み合わせにより、F0 分布のマルチスケール可視化を実現する。本論文では、1,000 人の歌唱者による同曲異唱を可視化し、被験者実験による可視化パラメータの最適化と有効性評価を行った結果を報告する。

1 はじめに

大規模データの分布や傾向を理解するための情報可視化技術は従来から活発に研究されている。近年では単に新しいデータ描画手段を開発するだけでなく、可視化への要求やタスクの定義、またそのタスクを実現するためのシステムデザインやインタラクションデザイン、さらには可視化結果の可読性や視認性の評価なども重要な研究課題である。

そこで本論文では、大規模データの情報可視化の事例として、同一楽曲に対する多数の歌唱データである「同曲異唱コンテンツ」を対象に、歌い方の分布や傾向を可視化するインタフェース SingDist Vis を提案する。人々が自身の歌った既存曲の歌唱 (動画) を動画共有サービスやソーシャルカラオケサービス¹へ投稿した同曲異唱コンテンツは、オリジナル曲の鑑賞の楽しみ方を拡張し、新たな好みのコンテンツやアーティストの発見につながる [6]。さらに、歌唱における声質と歌い方はそれぞれ、歌手の個性や類似性を特徴づける重要な要素であることから [26]、同一曲に対する多様な声質と歌い方の大規模データの分析は、人の歌声知覚・生成や歌唱スタイル等特性の違いを分析・理解する上でも重要である。

本論文では、歌唱における歌い方に着目し、音高として基本周波数 (F0) 推移の傾向・分布を可視化する。従来、大規模な楽曲データを対象として、そこから 1 曲を選択するために楽曲単位で可視化するインタフェースは存在したが (例えば, [30, 6] 等)、曲の内容も含めた可視化は研究されてこなかった。例えば、1 曲分の単独歌唱において F0 推移は折れ線で可視化されることが多いが、多数の F0 を単純に重ね合わせてしまうと図 1 のように違いや傾向を観察できず、また 1 曲を選び出すことは難しい。

そこで SingDist Vis では、「Overview+Detail」[3]

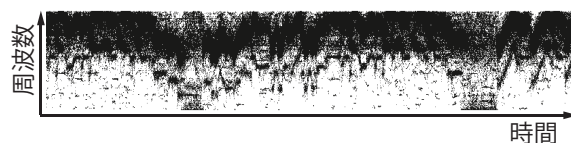


図 1. 多数の F0 を単純に折れ線集合で表示すると、過密による Visual Cluttering 問題が発生する。

という可視化デザインに沿った、以下の 2 画面の構成によるマルチスケールな可視化を実現する。

Overview: 大域的な F0 分布をヒートマップで概略的に表現する画面 (例: 曲の 1 番を全て表示)。

Detail: 局所的な F0 分布を細かな粒度で可視化するために、折れ線の集合で表現する画面。歌唱力の違いも色の違いで可視化する。

SingDist Vis の操作手順を図 2 に示す。ユーザはまず Overview 画面から、F0 推移及びその各時刻での分散の 2 種類のヒートマップ表示を用いて、異なる音高で発声している歌唱者がいるフレーズや、F0 が幅広く分散するフレーズが、楽曲の中でどのように存在しているかを観察する。続いて、特徴的な歌唱の発見等のため、Overview 画面で特定のフレーズを矩形領域として指定することで、Detail 画面でその F0 分布を折れ線集合として詳細に観察できる。さらに Detail で特定の歌唱を指定すると、その F0 推移を Overview 上で重畳させて観察できる。

本論文は我々自身の先行発表 [27] をインタフェース提案の観点からまとめ直している。歌唱力と F0 推移に関する知見から要求定義を検討し、パラメータ決定と評価のための被験者実験を追加している。

2 関連研究

同曲異唱を扱う研究事例として、Hamasaki *et al.* は、ニコニコ動画の歌ってみた動画を対象に、各歌唱の再生回数や男女度を可視化しながら、各歌唱を

* お茶の水女子大学

† 産業技術総合研究所

¹ ニコニコ動画, YouTube, Sing!カラオケ (Smule) 等。

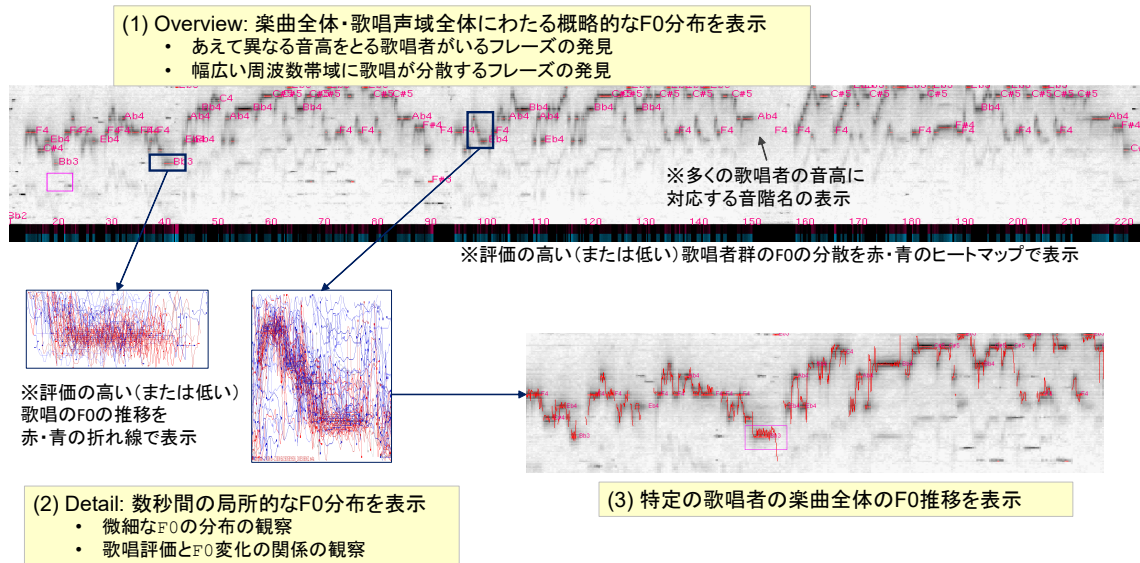


図 2. SingDistVis のスナップショットおよび操作手順. (1) Overview: 1 曲全体の F0 群と各時刻におけるその分散をヒートマップで可視化. (2) Detail: Overview で指定した矩形領域の F0 分布を, 歌唱力の違い (赤=高評価, 青=低評価) とともに, 折れ線の集合で可視化. (3) Detail で指定した F0 推移の Overview での重畳表示.

自動的につないで一つの楽曲としてリレー再生する Songrium RelayPlay を提案した [6]. 都築らは, 複数歌唱を同時再生して仮想的な合唱を制作できる Unisoner を提案し, 4,524 人の F0 を活用して F0 推定精度を向上させる技術, F0 分布を局所的にヒートマップで可視化する機能も提案した [28]. また近藤及び著者らは, 2,024 人の F0 に対し, ヒートマップを用いた大域的な可視化を提案した [25].

歌唱分析に F0 可視化を用いた研究事例として, 歌唱力向上等を目的として, 単一歌唱の F0 推移の可視化は多く対象とされてきた (例えば, [7, 29, 16, 13, 19] 等). ここで, F0 推移を音符単位にまとめ, 音高の楽譜からのずれを色の違いで可視化する研究もある [15, 24]. また, 歌唱スタイルの違いを可視化・識別することを目的として, 可視化の次元 (軸) を増やし, F0 と音量を 2 軸とした可視化 [11], F0 と F0 差分を 2 軸とした可視化 [9] が報告されている.

これらの研究に対して SingDistVis では, 数千人規模の F0 をマルチスケールで同時可視化する点で独自性を有する. 近藤及び著者らの研究 [25] を拡張し, Overview+Detail デザインを導入して, 概略と詳細のインタラクティブな可視化を可能とする.

3 大規模 F0 データの可視化方策

歌唱の F0 推移は時系列データとして扱うことができる. K 個の標本が, それぞれ L 個の時刻 t_l における実数値 $p_k(t_l)$ を有する場合, 以下の 3 種類の可視化手法の適用が可能である.

- 一方の座標軸を t_l , 他方の座標軸を $p_k(t_l)$ とした折れ線グラフ [22, 18, 23] や散布図 [23].

1. の折れ線や点群をその密度で表し, それを各画素の明度や色相に変換したヒートマップ.
- 一方の座標軸を t_l , 他方の座標軸を K 個の標本番号 k とした上で, $p_k(t_l)$ を各画素の明度・色相で表現したヒートマップ [8, 17], もしくは線分の太さ [10] で表現する可視化手法.

このうち 1. の折れ線グラフや散布図では, Visual Cluttering と呼ばれる視認性の低下が避けられない (図 1). 3. のヒートマップは, 色の識別能力の限界により数値を正確に読み取ることが難しく, 他方, 線分の太さの利用では描画に必要な画面サイズに限界があることから大規模データの可視化が難しい.

そこで, SingDistVis の Overview 画面には 2. の「密度のヒートマップ」を採用した. 一方で Detail 画面には, 微細な違いを視認しやすい 1. の「折れ線グラフ」を採用し, 表示対象となる折れ線の本数を制御することで Visual Clutter の影響を抑える.

従来, 時系列データの大局的及び局所的な時間変化に注視するマルチスケールな可視化としては, 時系列データを同じ視覚表現で拡大縮小するか [4, 20], 異なる時系列データを視覚表現を切り替えて組み合わせていた [5, 2]. これに対して SingDistVis では, 同一の時系列データ (F0) に対して, 異なる視覚表現を適用して歌唱分布を表現した点に特徴がある.

4 SingDistVis の実装

前章での議論に従い, 我々が開発した F0 分布の可視化手法の処理手順・操作手順を述べる. なお我々の先行発表 [27] では, より詳細な説明を含む.

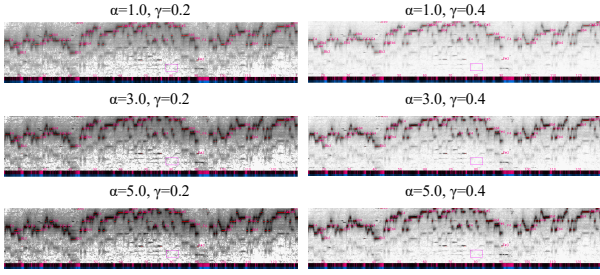


図 3. α および γ の調節に対する可視化結果の変化.

4.1 前提となるデータ：F0 と歌唱評価結果

本論文では、 K 個の歌唱で構成される集合 S において、その要素 s_k を以下のように定義する.

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

$$s_k = \{p_{k1}, p_{k2}, \dots, p_{kL}, e_k\} \quad (1)$$

ここで L は標本化された時刻の総数であり、 p_{kl} は k 番目の歌唱の l 番目の時刻における F0 の対数、 e_k は k 番目の歌唱への評価値 (5 章で説明) である. 無音部分の対数 F0 は、便宜上 $p_{kl} = 0$ とした.

このようなデータを対象として、「Overview」「Detail」の可視化ではそれぞれ以下のタスクを支援する.

Overview: 異なる音高で発声している歌唱のあるフレーズや、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、歌唱間の微小な 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.

4.2 Overview: ヒートマップでの大域表示とインタラクション

Overview では F0 推移の密度分布をヒートマップで表現する. 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. 時間方向は、歌唱の開始時刻および終了時刻を長方形領域の左右端に割り当てて N 個に分割する. また周波数領域の上限と下限を上下端にわりあて、これを M 個に分割する. 分割してできた各領域を通過する F0 の個数を集計することで、2次元ヒストグラムを生成し、これを $M \times N$ 画素のヒートマップ画像として扱う. 左から u 画素目、下から v 画素目の明度 I_{uv} は、その画素に存在する p_{kl} の個数 r_{uv}

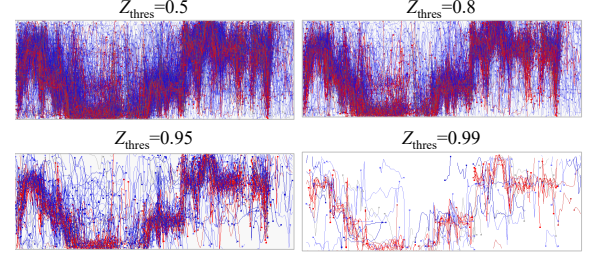


図 4. Z_{thres} の調節に対する折れ線の表示本数の変化.

から、以下のように求める.

$$I_{uv} = 1.0 - (\alpha r_{uv})^\gamma \quad (2)$$

ただし α および γ はユーザが調節可能なパラメータとし、その違いによる見た目の変化を図 3 に示す. α が大きいと r_{uv} が大きい部位での明度が低く (= 黒が濃く) なり、 γ が大きいと r_{uv} が大きい画素と小さい画素のコントラストが大きくなる.

Overview ではさらに、各時刻における F0 の分散もヒートマップで表示する. ここでは、歌唱評価 e_k の高い (低い) 歌唱群の F0 の分散を赤 (青) のヒートマップとした. これにより、「高評価な歌唱者が異なる歌い方をしているフレーズ」、「低評価な歌唱者が音高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、Detail で注目する矩形領域を決定する際の有力なヒントにもなる.

Overview には、時刻と F0 の範囲を示す矩形領域を重ねて表示する. ユーザは、マウスポインタの移動とともにこの矩形領域を移動させ、興味がある領域でクリックすることで、後述する Detail において局所的で詳細な可視化を観察できる.

4.3 Detail: 折れ線集合での局所表示とインタラクション

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が、前述した Visual Cluttering を抑制するために、同時に描画する折れ線の本数を制御する.

サンプリングおよびクラスタリングの 2 種類に基づく本数制御を実装したが [27]、本論文ではサンプリングによる折れ線の本数制御のみを説明し、後述の実験でもこれのみを用いる. これは、各々の折れ線 p_k 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 $z_k (0.0 \leq z \leq 1.0)$ を割り当て、以下を満たす場合のみ描画する.

$$\beta_{e_k} z_k > Z_{thres} \quad (3)$$

ここで、 β_{e_k} は歌唱評価 e_k 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. 例えば高 (低) 評価な p_k のみを表示したい場合は、その β_{e_k}

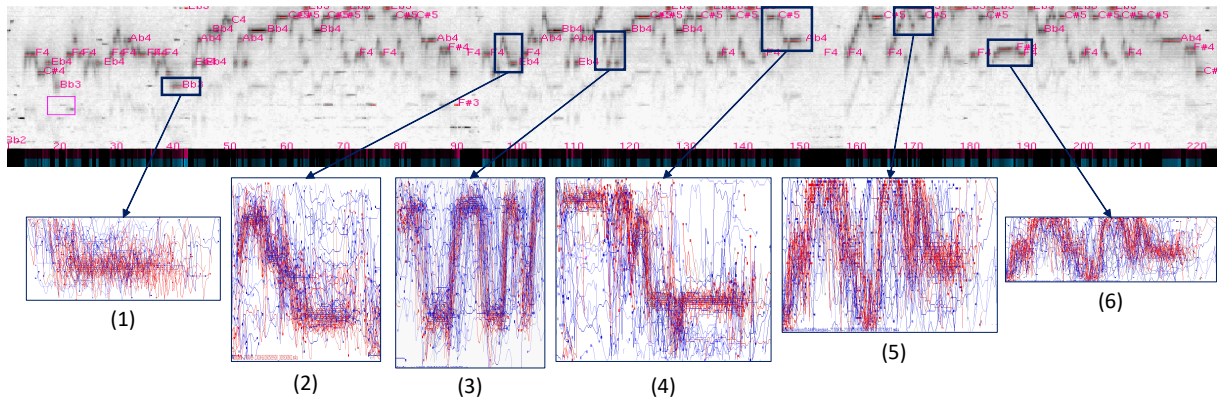


図 5. Overview と 6 箇所 Detail 例. Detail では、歌唱評価 e_k が高いほど赤く、低いほど青い折れ線で表示される。

が大きな値を取るように設定する。 $e_k = 1$ は歌っていない歌唱等に付与されたため、歌唱力が高い $e_k = 5$ と低い $e_k = 2$ の F0 が強調されるように、 $\beta_{e_k=2,5} = 1.0$, $\beta_{e_k=4} = 0.8$, $\beta_{e_k=1,3} = 0.5$ とした。

さらに、折れ線の色と不透明度を e_k に応じて変更 [27] することで、歌唱力の違いも可視化した (高いほど赤く、低いほど青い)。また、発声タイミングについても、F0 系列の開始と終了を強調した。

図 4 に、 Z_{thres} の値と折れ線の可視化結果を示す。

また、局所的な観察結果を大域表示と関連付けるために、Detail で特定の F0 推移をクリック操作で指定すると、Overview にも折れ線で重畳表示される。

5 SingDistVis の実行例

SingDistVis の実行結果を紹介する。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 (Let It Go Vocal Performances) [1] に収録された “Let It Go” を 231,278 人が歌った同曲異唱の無伴奏音響データから、1,000 人分を無作為抽出して適用した。これらの歌唱に対して pYIN [12] を用いて音響データから F0 を算出し、その相互相関最大化によって歌唱間の時間ずれを補正した。

Overview 可視化の画素数は $N = 1000$, $M = 480$ 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。

Detail で用いる歌唱力評価結果 e_k は、1 人の評価者が各歌唱を 5 段階のリッカート尺度に基づいて主観評価した結果とした。評価者は 12 年間ピアノの指導を受けた経験があり、対象楽曲 (英語原曲) を必要に応じて聴取しながら、各歌唱 1 コーラス目のサビ直前からそのサビ終了までを聴取して評価した。

5.1 Overview: Detail 表示箇所の探索

楽曲全体を表す Overview と、矩形領域を移動させた 6 箇所 Detail 例を図 5 に示す。図上部の Overview では、F0 分布ヒートマップが黒いほど多くの歌唱者が同一音高で発声していることを示す。

ユーザは、Overview の下端に表示される赤と青のヒートマップ (高・低評価の F0 分散) も参考にしながら、Detail を表示する領域を決定する。例えば、分散が大きい箇所は「歌い方が異なる」「難しいので多くの人が間違える」可能性があり、歌唱力が高くて異なった歌い方をしている歌唱者や、重点的に練習すべき箇所の発見等につながると思う。

5.2 Detail: 歌唱評価 e_k と F0 推移の詳細を観察

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) は赤い折れ線が太い帯を形成しており、高評価な歌唱であっても F0 に幅があった。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、何人かの低評価の歌唱者がそれ以上に大きく F0 を外していた。

図 5(3)(4)(5) は、いずれも音階が大きく上下するフレーズである。高評価な歌唱と低評価の歌唱とでは F0 の正確さが異なることが観察できた。また図 5(5)(6) では図 5(1) と同様に、フレーズの末尾において、高評価な歌唱者であっても F0 に幅があった。

5.3 Detail と Overview の重畳表示

図 5(1) に示す Detail から、特定の折れ線をクリックして、Overview にて重畳表示した例を図 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ており、楽曲全体にわたって正確な F0 で歌唱していた。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) のように、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った。このような「多数の歌唱者と異なる F0 推移をとりながらも高評価な歌唱者」の例をいくつか聴取したところ、歌唱技法 (ビブラート等) のスタイルや癖、タイミングのずれ、F0 誤推定、等の理由があった。

一方で低評価な歌唱の中には、多数の歌唱者 F0 推移に近い図 6(3) と、多数の歌唱者の F0 推移から大きく離れた図 6(4) が見られた。前者は F0 推移には表れない声質の問題 (声が細い、舌足らず) や、

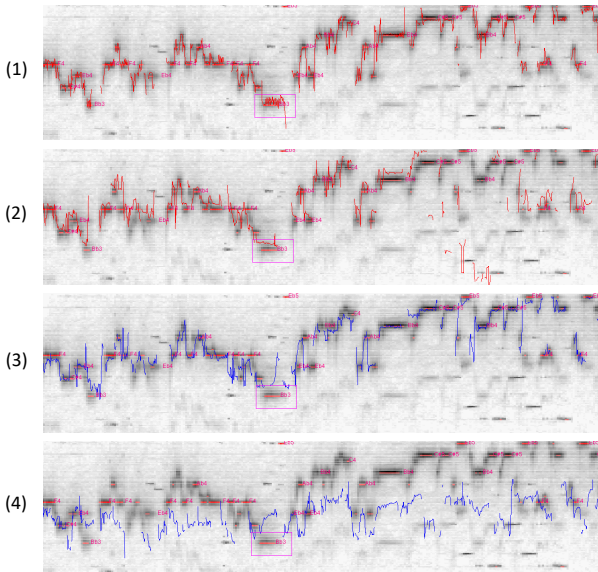


図 6. 特定歌唱者の F0 推移の例. (1)(2) 高評価と、(3)(4) 低評価の歌唱例. また F0 推移が多数の歌唱者と (1)(3) 類似するか、(2)(4) 異なる歌唱例.

声域の狭さ（高音部・低音部に難がある）等が原因で、低評価につながったと考えられる。

6 実験：被験者主観評価

SingDistVis の可視化パラメータの選定と、可視化デザインの有効性を被験者実験により評価する。条件統制のため、いずれの評価においても、被験者はインタフェース操作をせず、可視化結果の画像のみがディスプレイ上で呈示されて質問に回答した。

被験者は、ディスプレイによる閲覧が可能で色覚に不安がない、情報科学を専攻する学部学生 18 名であった。音楽経験は問わなかったが、実験前に可視化結果の見方を教示した。

6.1 可視化パラメータ値の選定

可視化結果の視認性向上には、適切なパラメータの設定が重要であることが知られている [23, 21, 14]。

本実験では、Overview におけるヒートマップの明度とコントラストに関する α 及び γ (式 (2)) と、Detail における折れ線の本数制御に関する Z_{thres} (式 (3)) について、パラメータを振って生成した複数画像と質問を被験者に呈示する。後述の質問 1~6 と画像の組合せをランダムに呈示した。

6.1.1 Overview：ヒートマップパラメータ

Overview については、後述する図 7 に示す α と γ の組み合わせで 20 通りの画像を生成し、左から右へ α が増加する順、下から上へ γ が増加する順で全てを同時に配置した 1 枚の画面を被験者に呈示

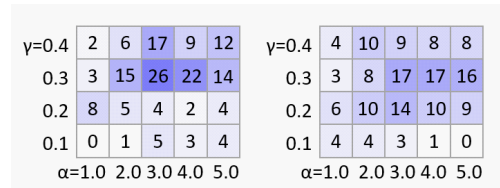


図 7. 質問 1(左) と質問 2(右) に対する回答分布。

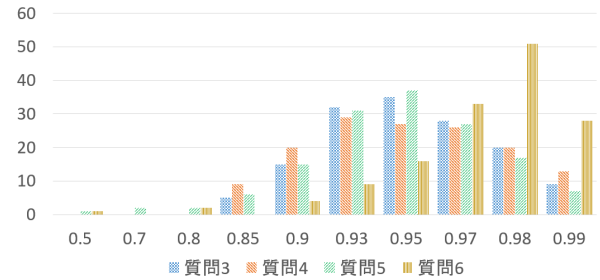


図 8. 質問 3~6 に対する回答分布。

した。全パラメータ値の可視化を同時に呈示するため、また曲中の場所に応じた影響の考察につなげるため、80 秒分のヒートマップを無作為に切り出した 3 箇所とした。3 箇所それぞれについて、20 通りの画像の中から以下の 2 つの質問に最も該当する画像を J_1 個選択してもらった ($J_1 = 3$)。

質問 1: 歌唱の全体表示として回答者が最も使いたい可視化はどれか。

質問 2: ズームアップして確認したい部分を選ぶのに向いた可視化はどれか。

3 箇所の可視化結果に対する被験者の選択結果の集計値を図 7 に示す。この結果から $\alpha = 3.0$, $\gamma = 0.3$ が良好なパラメータ値であると判断し、6.2 節での評価実験でもこの値を採用した。これらは全質問の全 3 箇所中、常に最多か 2 番目の回答数であった。

6.1.2 Detail：折れ線の本数制御パラメータ

Detail については、後述する図 8 に示す Z_{thres} の値によって、10 通りの画像を生成し、 5×2 の配置で左上から右下まで値が増加する順に配置した 1 枚の画面を被験者に呈示した。これを無作為に選んだ曲中の 4 箇所について、以下の 4 つの質問に最も該当する画像を J_2 個選択してもらった ($J_2 = 2$)。

質問 3: 高評価な歌唱者は正確な音高で歌唱しているか、高評価だけどもえて音高をずらしている歌唱者がいるか、正確な音高なのに低評価な歌唱者がいるか、といった点を発見しやすい可視化結果はどれか。

質問 4: 全体からずれた音高で歌唱している人が、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

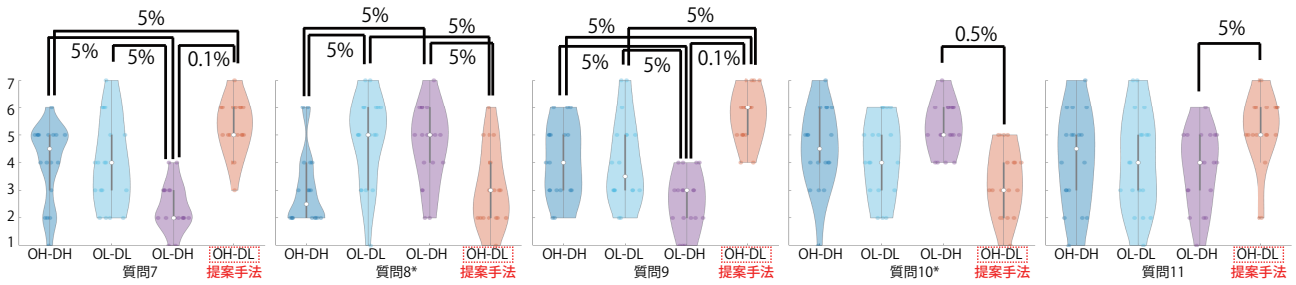


図 9. ポジティブな質問 7,9,11 とネガティブな質問 8*,10*に対する回答分布と仮説検定の結果（数字は有意水準）。

高がずれているのか、といった点を議論しやすい可視化結果はどれか。

質問 5: 歌唱評価結果と音高の正確さを比較しながら分析しやすい可視化はどれか。

質問 6: 注目すべき歌唱者を適切に発見できそうな可視化はどれか。

4 箇所の可視化結果に対する被験者の選択結果の集計値を図 8 に示す。6.2 節での評価実験では、全集計で最多であった $Z_{thres} = 0.95$ を採用した。

ここで、特定の歌唱を全体傾向と比較する際（質問 3~5）では、 $Z_{thres} = 0.9$ から 0.97 のいずれかが最多回答となり、選択箇所に応じて多少ばらつきがあった。また、特定の歌唱を発見する際（質問 6）では、より折れ線数の少ない $Z_{thres} = 0.98$ が好まれる傾向にあり、全 4 箇所ですべて最多の回答を得た。

6.2 可視化デザインの正当性評価

SingDistVis の Overview (O) と Detail (D) について、ヒートマップ (H) と折れ線 (L) を用いて、視覚表現の組合せ方と個別の評価を実施する。後述の質問 7~13 と画像の組合せをランダムに呈示した。

6.2.1 組合せ方の評価

4 通りの組合せ OH-DH, OL-DL, OL-DH, OH-DL (提案手法) を生成して比較した。被験者は、以下のポジティブもしくはネガティブ (*) な質問に、それぞれ 7 段階のリッカート尺度で回答した。

質問 7: 頻繁に使いたい

質問 8*: 不必要に複雑である

質問 9: 使いやすい

質問 10*: 使用するには技術的な支援が必要だ

質問 11: 大量の歌唱の同時可視化にふさわしい

回答の分布をバイオリンプロットで図 9 に示す。また Wilcoxon の順位和検定を行い、帰無仮説が棄却されたペアを有意水準とともに図中に示す。

この図から、提案手法 (OH-DL) は、ポジティブな質問 7,9,11 について他の組合せよりも数値が大きく、ネガティブな質問 8*,10* については数値が小さかった。例外として質問 8 については、OH-DH が高評価（最も低い中央値）であった。

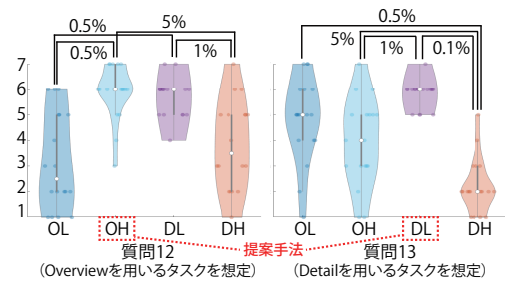


図 10. 質問 12~13 の回答分布と仮説検定結果。

6.2.2 個別の評価

OL, OH, DL, DH 4 種類の可視化を生成して比較した (OH と DL が提案手法)。

被験者は、以下の質問について 7 段階のリッカート尺度で回答した。

質問 12: 多くの歌手が同じ音高で歌唱している部分が発見しやすい

質問 13: 歌手によって歌い方の差が大きい部分が発見しやすい

回答の分布をバイオリンプロットで、図 10 に示す。また前項と同様、Wilcoxon の順位和検定の結果も示す。質問 12 は Overview を用いるタスクを想定した質問であり、提案手法 (OH) が高評価を得た。DL も高評価となったが、追加ヒアリングの結果、折れ線の密度で判断しやすいとの回答が多かった。

質問 13 は Detail を用いるタスクを想定した質問であり、これも提案手法 (DL) が高評価を得た。

7 まとめ・今後の課題

本報告では大規模な同曲異唱コンテンツの F0 分布を「Overview+Detail」という可視化デザインに基づいて探索的に可視化する手法を提案し、その実行例と被験者主観評価結果を示した。今後の課題として、6.1 節で選定した可視化パラメータ値の一般化 (異なるデータ数への適用等) や、インタフェース操作に関する評価実験を進めたい。またさらに、先行発表 [27] にも述べた通り、歌唱データ処理の改善、歌唱支援の実用シナリオに沿った可視化手法の拡張、歌唱以外のデータへの適用も検討していきたい。

参考文献

- [1] WELCOME TO STANFORD'S DAMP: Stanford Digital Archive of Mobile Performances, a repository of geo-tagged mobile performances to facilitate the research of amateur practices. <https://ccrma.stanford.edu/damp/>.
- [2] P. André, M. L. Wilson, A. Russell, D. A. Smith, A. Owens, and M. C. Schraefel. Continuum: Designing Timelines for Hierarchies, Relationships and Scale. In *Proc. ACM Symposium on User Interface Software and Technology (UIST 2007)*, pp. 101–110, 2007.
- [3] A. Cockburn, A. Karlson, and B. B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys*, 41(2):1–31, 2009.
- [4] E. Cuenca, A. Sallaberry, F. Y. Wang, and P. Poncet. MultiStream: A Multiresolution Streamgraph Approach to Explore Hierarchical Time Series. *IEEE Trans. on Visualization and Computer Graphics*, 24(12):3160–3173, 2018.
- [5] M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. Songle: A Web Service for Active Music Listening Improved by User Contributions. In *Proc. 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pp. 311–316, 2011.
- [6] M. Hamasaki, K. Ishida, T. Nakano, and M. Goto. Songrium RelayPlay: A Web-based Listening Interface for Continuously Playing User-generated Music Videos of the Same Song with Different Singers. In *Proc. International Computer Music Conference (ICMC 2020)*, pp. 1–4, 2020.
- [7] D. Hoppe, M. Sadakata, and P. Desain. Development of Real-time Visual Feedback Assistance in Singing Training: A Review. *Journal of computer assisted learning*, 22(12):308–316, 2006.
- [8] M. Imoto and T. Itoh. A 3D Visualization Technique for Large Scale Time-Varying Data. In *Proc. International Conference on Information Visualisation (IV10)*, pp. 17–22, 2010.
- [9] T. Kako, Y. Ohishi, H. Kameoka, K. Kashino, and K. Takeda. Automatic Identification for Singing Style Based on Sung Melodic Contour Characterized in Phase Plane. In *Proc. 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pp. 393–398, 2009.
- [10] M. Krstajic, E. Bertini, and D. A. Keim. CloudLines: Compact Display of Event Episodes in Multiple Time-Series. *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2432–2439, 2011.
- [11] K. W. E. Lin, H. Anderson, N. Agus, C. So, and S. Lui. Visualising Singing Style Under Common Musical Events Using Pitch-Dynamics Trajectories and Modified TRACCLUS Clustering. In *Proc. 13th International Conference on Machine Learning and Applications (ICMLA'14)*, pp. 237–242, 2014.
- [12] M. Matthias and S. Dixon. pYIN: A fundamental frequency estimator using probabilistic threshold distributions. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, pp. 659–663, 2014.
- [13] O. Mayor, J. Bonada, and A. Loscos. Performance Analysis and Scoring of the Singing Voice. In *Proc. AES 35th International Conference*, pp. 1–7, 2009.
- [14] L. Micalef, G. Palmas, A. Oulasvirta, and T. Weinkauff. Towards Perceptual Optimization of the Visual Design of Scatterplots. *IEEE Trans. on Visualization and Computer Graphics*, 23(6):1588–1599, 2017.
- [15] F. Moschos, A. Georgaki, and G. Kouroupetroglou. FONASKEIN: An Interactive Software Application for the Practice of the Singing Voice. In *Proc. Sound and Music Computing Conference (SMC) 2016*, pp. 326–331, 2016.
- [16] T. Nakano, M. Goto, and Y. Hiraga. MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data. In *Proc. IEEE International Symposium on Multimedia (ISM 2007) Workshops*, pp. 75–76, 2007.
- [17] G. Oliveira, J. Comba, R. Torchelsen, M. Padilha, and C. Silva. Visualizing Running Races through the Multivariate Time-Series of Multiple Runners. In *Proc. Conference on Graphics, Patterns and Images (SIBGRAPI 2018)*, pp. 99–106, 2013.
- [18] C. Perin, F. Vernier, and J.-D. Fekete. Interactive Horizon Graphs: Improving the Compact Visualization of Multiple Time Series. In *Proc. ACM CHI Conference on Human Factors in Computing Systems (CHI '13)*, pp. 3217–3226, 2013.
- [19] M. Shiraishi, K. Ogasawara, and T. Kitahara. HamoKara: A System that Enables Amateur Singers to Practice Backing Vocals for Karaoke. *Journal of Information Processing*, 27:683–692, 2019.
- [20] K. Sjölander and J. Beskow. Wavesurfer - An Open Source Speech Tool. In *Proc. International Conference on Spoken Language Processing (ICSLP 2000)*, pp. 464–467, 2000.
- [21] S. Smart and D. A. Szafrir. Measuring the Separability of Shape, Size, and Color in Scatterplots. In *Proc. ACM CHI Conference on Human Factors in Computing Systems (CHI '19)*, pp. 669:1–14.
- [22] Y. Uchida and T. Itoh. A Visualization and Level-of-Detail Control Technique for Large Scale Time Series Data. In *Proc. 13th International Conference on Information Visualisation (IV09)*, pp. 80–85, 2009.
- [23] Y. Wang, F. Han, L. Zhu, O. Deussen, and 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):1141–1154, 2018.
- [24] C. Weiß, S. J. Schlecht, S. Rosenzweig, and M. Müller. Towards Measuring Intonation Quality of Choir Recordings: A Case Study on Bruckner’s Locus Iste. In *Proc. 20th International Society for Music Information Retrieval Conference (ISMIR 2019)*, pp. 276–283, 2019.
- [25] 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 (インタラクティブ発表), pp. 745–750, 2021.
- [26] 齋藤 毅, 榊原 健一. 物真似歌唱の音響特徴とその知覚への影響の調査. 日本音響学会 秋季研究発表会 講演論文集, 2011, pp. 571–574, 2011.
- [27] 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分布のマルチスケール可視化. 情報処理学会 研究報告 音楽情報科学 (MUS), 2021–MUS–132–5, pp. 1–8, 2021.
- [28] 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):2370–2383, 2015.
- [29] 平井 重行, 片寄 晴弘, 井口 征士. 歌の調子外れに対する治療支援システム. 電子情報通信学会論文誌, J84-D-II(9):1933–1941, 2001.
- [30] 魚田 知美, 伊藤 貴之. GRAPE: グラデーション画像による ポータブルビジュアルプレイリスト. 第 41 回可視化情報シンポジウム講演論文集, pp. 299–304, 2013.

未来ビジョン

本論文が提案する SingDistVis は、大規模な同曲異唱データの中から特徴的な歌唱を対話的に発見する新しいインタフェースである。本論文ではその研究の初期段階として、インタフェースの根幹となる可視化手法を示し、ユーザ評価実験によりその可視化デザインの妥当性を検証した。一方で、現状では、同曲異唱データにおける F0 の分布を概観し、特徴的な F0 推移を有する歌唱を発見するための汎用的な手法を開発した段階であり、今後は歌唱支援のための具体的なタスクを想定した拡張が重要であると考えられる。

本研究が目指す未来の一つとして、多数の歌唱者の中から自分がお手本としたい歌唱者・歌唱技法を探すことで多様な歌唱技法を理解し、その結果として発見した歌唱にならうことで歌唱力を向上させる、という一連のタスクを支援するインタフェースとしての拡張が考えら

れる。あるいは、同一楽曲を反復練習した過程を可視化することで、反復練習によってどのフレーズにおいて歌唱者がどの程度上達したか、あるいはどのフレーズにおいてなかなか上達していないか、といった点を理解することも可能となる。

さらに大きなビジョンとして、本研究の概念を拡張することで、歌唱に限らず多くのパフォーマンス（例えば楽器演奏、ダンス、スポーツなど）において、お手本としたい「自分だけの先生」の発見、あるいは各自の上達過程の観察などを可視化によって支援するフレームワークを提案し、「レッスン支援可視化 (LessonVis)」という研究ジャンルを確立したい。多様なパフォーマンスに取り組むユーザが各自のパフォーマンスをビジュアルに理解し、着実な上達につなげることで、パフォーマンスの楽しみ・夢・充実感をさらに向上させる、という環境を提供することが本研究の未来ビジョンである。