情報処理学会論文誌 Vol.64 No.10 1423–1437 (Oct. 2023)

ポピュラー音楽の模倣歌唱における歌唱テクニック分析と楽譜情報との対応付け

山本 雄也[1,a)] 中野 倫靖[2,b)] 後藤 真孝[2,c)] 寺澤 洋子[1,d)]

受付日 2022年11月21日, 採録日 2023年7月5日

概要：歌手は楽譜に合わせて歌うだけでなく，その音高や音色に変動を加えることがある．本論文は，これらを「歌唱テクニック」とし，ポピュラー音楽の中でも J-POP を対象に，歌手によって歌唱テクニックがどのくらいの頻度でどのように生起するか，そして楽曲のどこで生起するか，その傾向を分析することを目的とする．そこで本論文では，J-POP のプロ歌手 24 名（男女各 12 名）の歌い方を別のプロ歌手 14 名（男女各 7 名）が学術目的で模倣した歌声データベース「AIST-SIDB」に含まれる 48 歌唱の 13 種類の歌唱テクニックを対象として，歌唱テクニックとメロディの持つ音楽要素との関係性を分析した．具体的には，歌唱テクニックの生起頻度と，歌唱テクニックの 1 つであるビブラートに関してはそのパラメータ（深さと速さ）を分析した．さらに，歌唱テクニックの生起位置を楽譜情報と対応付けて，各歌唱テクニックと「歌詞の音素」,「音高」,「音高差」,「音長」,「フレーズ内における位置」との関係を分析し，またビブラートパラメータと「音高」および「ビブラート長」との相関を分析した．

キーワード：歌声情報処理，歌唱テクニック，模倣歌唱，ポピュラー音楽，ビブラート，J-POP

Singing Technique Analysis with Correspondence to Musical Score on Imitative Singing of Popular Music

Yuya Yamamoto[1,a)] Tomoyasu Nakano[2,b)] Masataka Goto[2,c)] Hiroko Terasawa[1,d)]

Received: November 21, 2022, Accepted: July 5, 2023

Abstract: Singers often add fluctuations in pitch and timbre to their vocal performance beyond just singing along with musical scores. This paper investigates them as “singing techniques” and aims to analyze how often and how they are used by singers, and where they are used in a song with a focus on J-POP of popular music. In this paper, we used AIST-SIDB, a singing database of 48 singing voices in which 14 professional singers (7 male and 7 female singers) imitated singing techniques of 24 professional singers (12 male and 12 female singers) of J-POP for academic purposes. We then analyzed 13 types of singing techniques included in the database in relation to musical scores. Specifically, we analyzed the frequency of occurrence of each singing technique as well as parameters (rate and extent) of vibrato, one of the singing techniques. We mapped the timings of the singing techniques to the musical scores, and analyzed the relationships of each singing technique to the “lyric phoneme,” “pitch,” “pitch interval,” “duration,” and “location within a phrase.” We also analyzed the correlation of vibrato parameters with “pitch” and “vibrato duration.”

Keywords: singing information processing, singing technique, imitative singing, popular music, vibrato, J-POP

1. はじめに

本論文では，ポピュラー音楽において重要な役割を担う歌声 [1] を対象とし，特に「歌唱テクニック」の分析を行

1 筑波大学人と音の情報学研究室
University of Tsukuba, Lab. for Sound and People with Computing (LSPC), Tsukuba, Ibaraki 305–8577, Japan
2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305–8568, Japan
a) s2130507@s.tsukuba.ac.jp
b) t.nakano@aist.go.jp
c) m.goto@aist.go.jp
d) terasawa@slis.tsukuba.ac.jp

© 2023 Information Processing Society of Japan

う．ここで歌唱テクニックとは，(1) 歌唱者の持つ動機やメッセージ等の表現意図，(2) 歌唱スタイル，そして (3) 歌唱者が楽譜や歌詞から汲み取った意図，の 3 つを具体化する方法論として定義する[*1]．歌唱者が歌唱テクニックを実践することで，観測可能な音響信号として具象化することになるが，それは具体的には，歌声における声質（声道共鳴特性，声帯音源特性）と歌い方（歌い回し）の制御によって実現される．歌声における声質と歌い方は，歌手の個性や類似性を特徴付ける重要な要素としても知られていることから [3]，その両者の精緻な制御としての歌唱テクニックを分析することは，歌手の個性や多様性の解明につながる．さらに歌唱テクニックに関する特徴の解明は，歌唱力評価 [4], [5]，歌唱練習支援 [4], [6]，歌声変換・加工 [5]，歌声検索 [4], [7]，歌声合成 [4], [5], [8], [9], [10] 等への幅広い応用が期待できる．

従来, ビブラート等の一部の歌唱テクニックについては, その特徴が分析されたり（たとえば, 文献 [11], [12], [13], [14]），パラメータ化に関する検討やその応用が研究されたりしてきた（たとえば，文献 [15], [16], [17], [18], [19], [20]）．それに対して本論文では，ポピュラー音楽において用いられる歌唱テクニックを幅広く調査したうえで，利用可能な歌声データベースに含まれるテクニックを対象として，扱うテクニックの数を増やして分析することを目指す．またその分析方法としては従来行われたパラメータ化（特徴）のみではなく，楽譜情報と対応付けて，それらがどこで（生起箇所）どれくらい（頻度）使われているのかも分析することで，文脈を考慮した歌唱テクニックの出現傾向の把握や，歌声情報処理システムへの応用へつなげる．

本論文ではポピュラー音楽の一ジャンルとして J-POP（日本で制作されて流行するポピュラー音楽）における歌唱を対象に，以下の 3 つの分析を行った．

(1) **歌唱テクニックの頻度・種類の集計**：楽曲ごとにどの歌唱テクニックが，どの程度使われているかを集計し，歌手ごとにその傾向の実態把握を行う．
(2) **楽譜情報を用いた歌唱テクニックの生起箇所の分析**：歌唱に基づいて書き起こした音高系列から，楽譜情報（音高・音程・音長・母音の音素・フレーズ内相対位置）との共起を集計し，歌唱テクニックと楽譜の関係を分析する．
(3) **歌手ごとのビブラートパラメータの分析およびその生起箇所の関係の分析**：ポピュラー音楽の歌唱テクニックにおいて頻出して重要性が高いビブラートを対象に，そのパラメータを歌手ごとに分析する．また，生起箇所とビブラートパラメータとの相関分析を行い，ビブラートの歌い方と音楽的文脈の関係を分析する．

[*1] この定義は，Umbert ら [2] による，歌声合成システムにおける歌唱表現コントロールのフレームワーク（文献 [2] の図 1）を参考にした．

上記分析には，J-POP の特性を反映するために，商業音楽においてよく知られたプロ歌手の歌い方を対象として，それを別のプロ歌手が学術目的で模倣した歌声を用いる．商業音楽は一般的には伴奏付きの歌唱であり，音源分離適用後もボーカルエフェクト等に由来するノイズやメインボーカルに重なるバックコーラスのパート等が残留するため，音響特徴の分析が困難な場合がある．そのため，商業音楽そのものを対象とはしない．プロとして活動している歌手であれば，模倣により商業音楽のプロ歌手の歌い方の特性を反映できる可能性が高い．また，模倣歌唱を用いる利点として，歌手の個性が歌唱テクニックとどう関係して知覚されうるのか，また別の歌手によってそれがどう生成されうるのか，という観点からの分析につながる可能性がある．このような模倣歌唱を用いる分析方策は，人の歌声知覚・生成を理解するうえでも重要な取組みである [3]．

2. 関連研究

本章では，歌唱テクニックの分析やデータセット，楽譜と歌唱テクニックの関連性に関連する研究や文献を分類してまとめる．結びには，それらの先行事例をふまえ，本論文の位置付けについて述べる．

2.1 歌唱テクニックの特徴分析と応用

歌唱テクニックとしては，従来，声楽歌唱を対象としたビブラートの音響特性および知覚特性の分析（たとえば，文献 [11], [13], [14], [15], [21]）が行われてきた．ビブラートとは，主に音を伸ばすときに周期的に F0（や音量）を変化させる（揺らす）歌唱テクニックであり，熟達した歌唱者が頻繁に用いる重要なテクニックである．ビブラートは，深さ（extent：ビブラート区間の音高方向の変動幅）と速さ（rate：毎秒に生じる揺らぎの回数）の 2 つのパラメータの観点から分析・制御されることが多い．

中野らはそれらのビブラートパラメータを F0 軌跡から抽出する手法を提案して歌唱力評価に応用し [16]，Gupta らもそれらパラメータの歌唱力評価への有効性を分析した [22], [23]．右田らは，さらに時間変動やビブラート区間長も考慮してビブラートを制御することで，歌手の個性を反映したビブラート歌声を合成した [24]．また Yang らは，ビブラート形状（sinusoid similarity と envelope）も考慮し，かつポルタメントに関するパラメータ化も行って京劇（Beijing Opera）の歌唱スタイルの違いを分析した [20]．Ikemiya らは，ビブラート以外のテクニックとして，F0 軌跡の形状で特徴付けられるこぶし・グリスアップ・フォールダウンも対象とし，それらをパラメータ化して既存楽曲の歌い方を変更する手法を提案した [18]．Miyazaki らは，VOCALOID を用いて人工的に生成させたビブラートを用いて，ビブラートパラメータと自然性の関係の主観評価を行った [25]．

© 2023 Information Processing Society of Japan

F0 軌跡以外を対象として，声質の制御による歌唱テクニックが存在する．Lee らは，ポピュラー音楽やロック音楽等の訓練された歌手のミックスボイスを含む声区を対象として，ハイスピードカメラ画像や EGG 等に基づく分析に加え，高調波構造における振幅の違いを示した [26]．ミックスボイスとは，地声（胸声）と裏声（ファルセット）の中間の声区として定義され，熟達した歌手によって用いられる歌唱テクニックの 1 つである．その他，裏声等の声区識別では，そのような振幅スペクトルの包絡形状を含む特徴量が提案されている [27], [28], [29]．

これらの研究に対する本論文の差分は，ロングトーンやスケール，アルペジオ等の実験刺激的な音形ではなく，実際の楽曲で用いられている歌唱テクニックを対象に扱っている点である．また，実際の楽曲を扱った Ikemiya ら [18] は 4 種類を扱っていたが，第 1 著者による文献調査 [30] では少なくとも 32 種類あげられたことから，本研究では 13 種類として対象数を増やした結果を報告する．

2.2 歌唱テクニックに関するデータセット

歌唱テクニックに関するデータセットも存在し，Wilkins らは，Breathy 発声や Fry 発声等，17 種類の異なる歌唱テクニックで歌唱されたデータセット（VocalSet）を構築し，歌声波形を入力とした深層学習に基づく識別手法を提案した [31]．山本らは，VocalSet を用いて MFCC およびビブラートパラメータを用いた識別と，深層学習に基づくテクニックの識別性能を比較した [32]．Kim らは，ウィスパーボイスやシャウト等の歌唱テクニックを含む，70 種類のタグでアノテーションされたデータセット（KVT dataset）を構築し，その特性を分析した [33]．Kalbag らは，ヘヴィメタルの楽曲を対象に 3 種類のスクリーム（High，Middle，Low）の時刻付きアノテーションを付与したデータセット（MVD Dataset）を構築し，通常歌唱，非歌唱部分を含むスクリーム識別問題に取り組んだ [34]．Yamamoto らは，既存歌手 42 名による各 4 曲の計 168 曲からなる J-POP の既存楽曲の CD 音源に対して歌唱テクニックのアノテーションを付与したデータセット（COSIAN）を構築し，9 種類の歌唱テクニックの検出に取り組んだ [35]．

VocalSet [31] はアルペジオ，スケール等の簡易なパターンのみで歌唱されており，実際のポピュラー楽曲における歌唱の分析に適さない可能性がある．また，KVT dataset [33] は歌唱テクニックに特化して構築されたデータセットではない．MVD dataset [34] はヘヴィメタル楽曲のスクリームのみを対象としているため，本研究の主旨とは異なる．COSIAN [35] は，J-POP を対象にしており，かつ多種類の歌唱テクニックのアノテーションが付与されているが，楽譜情報のアノテーションはなく，その対応付けを分析することは不可能である．加えて，その音源は伴奏付き歌唱であり，1 章で述べた既存楽曲そのものを用いることへの懸念がある．

本論文はポピュラー楽曲の歌唱中に現れる歌唱テクニックを楽譜情報として分析することが目的であり，そのためには以下の要件を満たすデータセットが望ましい．

- スケールやアルペジオ等実験刺激的な音形ではなく，楽曲のメロディを歌唱していること
- アカペラかつ高品質録音環境下で収録された歌声であること（すなわち，音響特徴を分析する際に，伴奏音等の影響を受けないこと）
- 歌唱メロディの楽譜情報と歌唱テクニックの出現区間，両方のアノテーションが付与されていること

先行事例として前述したデータセットには，これらの要件をすべて満たすものはない．

2.3 歌い方に関する特徴分析と応用

歌声には，意図的に制御できる歌唱テクニックとは別に，身体的な制約による動的変動成分が存在する．Saitou らは，制動 2 次系伝達関数を用いた分析を行い，ビブラート，オーバーシュート，プレパレーション，微細変動の 4 種類の動的変動成分と歌声の自然性の関係を示し，それらを制御することで自然な歌声合成を可能とした [15]．また Ohishi らは，歌声からそのような変動成分を音符情報から分離する確率モデルを提案し，変動成分を変更して歌い方を変えた再合成を可能とした [36]．

このような動的変動成分を含めた，歌い方全般に着目した研究として，Kako らは基本周波数（F0）とその微分（ΔF0，ΔΔF0）で描かれる相平面上の軌跡（ビブラート等）を GMM でモデル化することで，歌唱スタイル（クラシック，ポピュラー音楽，アマチュア）を識別する手法を提案した [37]．また Ardaillon らは，歌声の F0 軌跡における Attack/Release，Sustain，Transition の各状態を，B-スプライン曲線でモデル化して歌声合成へ応用した [19]．

歌唱スタイルやジャンル，歌唱力や経験等を分析するうえで，歌い方の違いが議論される事例もある．Panteli らは，ワールドミュージックを対象として，歌唱の F0 軌跡をクラスタリングすることで歌唱スタイルの類似性を分析し [38]，Abeßer らは，その特徴量と畳み込みニューラルネットワーク（CNN）による識別性能を比較した [39]．Repetto らは，中国の歌劇（Jingju）における 2 種類の歌唱スタイルの違いをビブラートパラメータやスペクトルセントロイド等を用いて比較した [40]．また Miryala らは，インド歌謡の Raga における歌唱表現分析として音符間の F0 軌跡の遷移（glide）に着目した [41]．その他，Mauch らは，合唱におけるチューニングが数十秒にわたって変動する（ずれる）intonation（pitch）drift と呼ばれる現象に着目し，その変動と音高の正確さを定量化して，歌唱力や経験による違いを分析した [42]．

本研究では，特に重要なビブラートを対象に，そのパラ

© 2023 Information Processing Society of Japan

メータである rate と extent について，プロの歌手におけるその分布を示し，楽譜情報と対応付けて分析する．

2.4 歌唱と楽譜情報を対応付けた分析と応用

歌唱テクニックや歌い方に関して，楽譜における文脈をともなって分析した研究がある．Nichols らは，歌詞情報と楽譜情報とを対応付け，強勢位置や母音と子音の違い等の言語的特徴と，小節内位置や長さ等の音楽的特徴の関係を分析した [43]．また，Arai らは，歌声の「グルーブ感」に着目し，母音と子音の発音開始時刻のずれ分布等を示した [44]．その他，音響分析をともなわないが，中里は J-POP における基準音から意図をもって外して歌ったこぶしやビブラートを対象として，楽譜上の出現位置と種類について考察した [45]．これらに対して本研究では，歌唱テクニックの楽譜上の出現位置と種類について定量的な分析と考察を行う．

楽譜情報による歌い方への文脈的な影響が明らかになれば，文脈情報を用いた応用につながる．たとえば，歌声合成への応用が可能である．楽譜情報と歌唱テクニックの関係を明示的に考慮して歌声合成を行う研究には以下がある．山田らは，ビブラートと文脈（小節内位置等）を考慮した隠れマルコフモデル（HMM）による歌声合成を提案した [46]．また Bonada ら [47] は，音符系列からビブラートや音高遷移等のパラメータを推定するモデルを提案し，Hono ら [48], [49] は歌声合成において，ビブラートパラメータや楽譜からの時間ずれを明示的にモデル化した．

本研究で，より多様な歌唱テクニックと楽譜情報の対応付けを行って分析することで，その結果を歌声合成技術へ今後応用することが考えられる．

3. 対象とする歌唱データと歌唱テクニック

本論文では既存の歌唱を模倣した歌唱を用いるため，そのデータについて説明する．また，対象とする歌唱テクニックの選出方法とラベル付けについて述べる．

3.1 歌唱データ

分析用の歌声には，AIST Singing Imitation Database（AIST-SIDB）の日本語歌唱を用いる．AIST-SIDB は，学術目的で模倣した歌声を収録して産業技術総合研究所（AIST）が構築した非公開の歌声データベースである．これは，J-POP における有名歌手の歌い方を，別のプロ歌手に対し模倣するように指示して歌唱してもらったアカペラの歌声で構築されている．ここで，AIST-SIDB の歌手は，その得意ジャンルや歌い方が歌唱対象曲とマッチするように選出された．

本論文は J-POP において観測できる歌唱テクニックの分析を研究目的としている．したがって，AIST-SIDB のように模倣歌唱であっても，プロとして活動している歌手の歌声であれば，商業音楽のプロ歌手の歌い方の特性が反映されている可能性が高いと考え，これを用いた．模倣する歌い方の対象は，**表 1** に示す 24 名の歌手である．表に示すように，歌手（楽曲）ごとに 2 名の異なるプロ歌手が真似て歌唱した歌声を収録しており，48 歌唱で構成される．各曲のキーが歌手の声域とマッチするように変更し，1 番のみをアカペラ歌唱として収録した．以降，本論文では便宜上，対象となった楽曲を「原曲」，原曲を歌ったプロ歌手を「原曲歌手」，実際に楽曲を歌唱した AIST-SIDB の歌手を「模倣歌手」と呼んで区別する．また，2 名の模倣歌手を区別して言及する場合には，文中では模倣歌手 1，図中では，{歌手名} {No.}（例：aiko 01）のように記述する．

表 1 AIST-SIDB の歌手の一覧（YUKI の motto の F02 の横の (−1) はキーを 1 半音下げて歌唱したことを示す）

Table 1 List of singers for AIST-SIDB. The (−1) following F02 for YUKI's motto indicates the key control for a semitone lower.

原曲歌手名（ID）	原曲名	性別	模倣歌手 1	模倣歌手 2
玉置浩二	出逢い	男	M03	M04
小田和正	キラキラ	男	M02	M06
GACKT	ありったけの愛で	男	M01	M05
桑田佳祐	勝手にシンドバット	男	M06	M07
チバユウスケ	カナリヤ鳴く空	男	M05	M01
西川貴教	Heat Capacity	男	M05	M01
hyde	Lies and Truth	男	M03	M04
平井堅	瞳を閉じて	男	M01	M05
福山雅治	桜坂	男	M04	M03
槇原敬之	桃	男	M04	M03
森山直太朗	さくら（独唱）	男	M02	M06
山崎まさよし	未完成	男	M06	M07
aiko	ボーイフレンド	女	F01	F06
絢香	三日月	女	F05	F02
宇多田ヒカル	Can You Keep A Secret?	女	F03	F04
鬼束ちひろ	月光	女	F04	F07
倖田來未	夢のうた	女	F06	F01
小柳ゆき	愛情	女	F04	F07
chara	大切をきずくもの	女	F02	F05
浜崎あゆみ	seasons	女	F06	F01
一青窈	ハナミズキ	女	F05	F02
平原綾香	明日	女	F03	F04
松浦亜弥	♡ 桃色片思い ♡	女	F01	F06
YUKI	motto	女	F02(−1)	F05

AIST-SIDB の模倣歌手は，日本語が母国語のプロとして活動している 14 名（男女各 7 名）であり，1 名が 2 曲もしくは 4 曲を歌唱した（模倣歌手の得意ジャンルが対象曲とマッチするように選出）．また，収録に用いたマイクは NEUMANN U87 Ai，ヘッドアンプには AMEK System 9098 Dual Mic Amplifier，コンプレッサには TUBE-TECH CL 1B を用いた．マイクと歌手の口元の距離は最低 60 mm（ふかれ防止フィルタからマイクの距離）として，録音に特化したスタジオで収録した．

3.2 楽譜情報

本論文では歌唱テクニックとメロディの楽譜との関係性を分析するため，メロディに関する音楽要素である，音高・音程・音長・母音の音素・フレーズ内の位置を対象とする．

© 2023 Information Processing Society of Japan

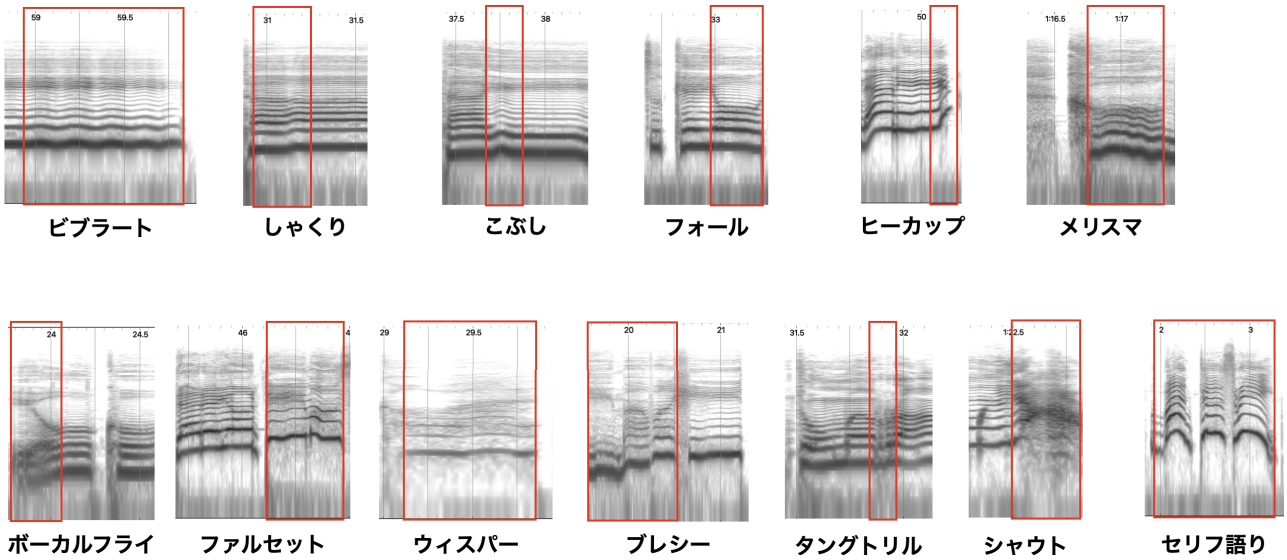

図 1　本論文で扱う歌唱テクニックの実例（スペクトログラム）．赤枠内を生起部分としてラベルを付与した

Fig. 1　Spectrograms of singing techniques. Annotated regions are surrounded by red bounding boxes.

そのためにまず，楽曲のメロディの楽譜情報（音高・発音時刻・音長）および歌詞を，プロの音楽家 1 名が，MusicXML の形式で原曲に対して採譜した．続けて，音楽要素のうち，音高・音程・音長は楽譜情報の音高から，母音の音素は歌詞から情報を取得した．フレーズは，歌唱を行う際のまとまりとして，本論文では楽曲の楽譜の全区間を 16 分休符より長い休符が生じる部分で分割した部分とした．

細かく変動する楽曲のテンポと同期するように採譜を行い*2，移調せずにすべて原音で表記した．また，シャッフルビートは 8 分音符や 16 分音符に置き換えず，すべて連符で記した．しゃくりやフォール等，歌唱テクニックやスタイル・クセに起因すると思われる音高変化は，基本的には採譜しなかった．ただし，音符を取り除くと別の曲に聞こえる等，その音符が明らかにメロディの構成要素であると判断した場合は含めた．

3.3　歌唱テクニック

分析対象の歌唱テクニックは，まず，ポピュラー楽曲で用いられている歌唱テクニックを可能な限り広く網羅できるよう，第 1 著者が既存の歌唱文献調査を行った [30]．その結果，音高制御に基づくもの（ビブラート等），音色・声質制御に基づくもの（ウィスパーボイス等），それ以外（シャウト等）に分類される，32 種類が得られた．続いて，前述した歌唱データに対して，この 32 種類に基づいてラベル付けを行い（方法は後述）．データ中に生起した歌唱テクニック 13 種類を本論文における対象とする．

*2　特に生演奏の伴奏が用いられる楽曲の場合，レコーディングの際に，クリック（メトロノーム）をまったく使用しないか，クリックのテンポを意図的に変動させている場合がある．

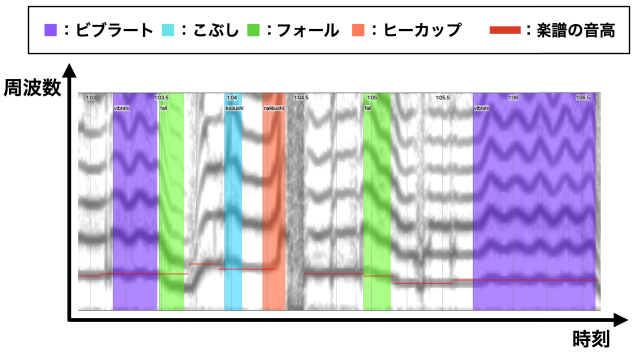

図 2　歌唱テクニックのラベル例（時刻および種類）

Fig. 2　The example of annotation (time stamps and types).

そのようにして決定した歌唱テクニックの実例と定義を，それぞれ**図 1** と**表 2** に示す*3．文献調査の結果，そのうちの 10 種類のテクニックは異なる名称で呼ばれることがあった．確認できた名称を表 2 の第 4 列に示す．以降，各テクニックの名称には第 1 列を採用して一意に扱う．

3.4　歌唱テクニックラベルのアノテーション

歌唱テクニックは種類と区間（開始時刻および終了時刻）をラベルとし，第 1 著者（バンドボーカル経験 9 年，合唱経験 6 年）がアノテーションを行った．Sonic Visualiser [50] を用い，歌声のスペクトログラムとそのピーク（F0 に相当する情報が確認できる）を可視化しながら作業した．**図 2** にラベル例を示す．

*3　メリスマに関しては，装飾音であったりフレーズとして音符が割り振られたりすることもあり，つまり（楽譜にない）歌唱テクニックとは区別される場合もあるが，メリスマを歌唱テクニックとして言及している事例 [38] があったため，今回の分析対象に含めた．

© 2023 Information Processing Society of Japan

表 2　本論文で扱う歌唱テクニックとその定義・名称ゆれ，および類似テクニック名
Table 2　Definitions, synonyms, and comparable techniques for the singing techniques.

歌唱テクニック名	変動する要素	変動の様子	名称ゆれ	類似するテクニック
ビブラート	音高・音量	揺らすように歌い，周期振動をつける	シェイク	トレモロ，トリル
しゃくり	音高	音高を上に連続変化させる	グリッサンド，ポルタメント，しゃくり上げ，スクープ	
こぶし	音高	U 字/逆 U 字状に音高を連続変化させる	ベンド，トレモロ	
フォール	音高	音高を下に連続変化させる	ドロップ，しゃくり下げ	
ヒーカップ	音高・音色	瞬時的に裏声または喉を締めた歌声を出す	なき，なき節，ボーカルブレイク	ヨーデル
メリスマ	音高	1 音節に複数の音高を割り当てる	フェイク	
ボーカルフライ	声質	ざらついた音を出す	エッジボイス，きしみ声	グロウル
ファルセット	声質	裏声区で歌唱する	ヘッドボイス，裏声	
ブレシー	声質	呼息音を混入する		
ウィスパーボイス	声質	囁くように歌唱する		
シャウト	−	叫ぶ		
セリフ語り	−	話すように歌う		ラップ
タングトリル	−	巻き舌を入れる	タングロール，巻き舌	リップロール

4.　分析

各歌唱テクニックの模倣歌手ごとの頻度を集計することで，テクニックの出現分布として扱うことができることから，大域的な歌手の個性を考察できる．また従来研究と同様に，歌唱テクニックの音響パラメータを算出することで，より詳細な個性に関する考察が可能となる．

最後に，頻度とパラメータを楽譜と対応付けることで，歌唱テクニックにおける文脈の影響について分析できる．

4.1　頻度の分析

全 48 歌唱における各歌唱テクニックラベルの頻度として，全時間長 [s] と平均継続長を**表 3** に示す．また，各歌唱テクニックにおける継続時間長の分布を**図 3** に示す．これらは J-POP としての歌唱テクニックの継続長に関する特性を意味する．多くの歌唱テクニックラベルの継続長は 0.1 秒から 1 秒の間に集中していることが分かる．ただし，ウィスパーボイスやブレシー等の声質の制御によるテクニックや，セリフ語り等のテクニックは，1 秒以上の長い継続長を持つ場合があった．このような長い継続長を持つテクニックは，意図的に使ったり使わなかったりを選択できるものだけでなく，歌手個人の声質や音域，楽曲等の特性を要因として，意図せずに使う，もしくは使わざるを得ない箇所が含まれている可能性がある．ファルセットに関しては，メロディに含まれる局所的な跳躍音で用いる場合と，フレーズ全体で連続して用いる場合があったため，継続長の分布として広くなった．

一方，模倣歌手ごとの特徴を分析するために，全 48 歌唱における歌唱テクニックの頻度分布を**図 4** に，頻度分布間のコサイン類似度[*4]を**図 5** に示す．図 5 から，模倣歌手 1 と模倣歌手 2 の多くは類似度が高く，0.9 以上が 54.17% (13/24)，0.8 以上が 75% (18/24)，0.7 以上が 91.7% (22/24) であった．0.7 以上で最も類似度が低い原曲歌手は西川貴教の 0.71 であり，最も類似度が低い原曲歌手は小田和正の 0.5 であった．図 4 より，前者についてはボーカルフライに差があり，後者についてはブレシー・しゃくり・ビブラートに差があることが分かる．それぞれの歌唱における頻度分布において，異なる原曲歌手を模倣した歌唱間との類似度の平均は 0.70 であり，同じ原曲歌手を模倣した歌唱間との類似度の平均は 0.83 であった．このことを考慮すれば，今回のデータにおいてはコサイン類似度の観点，つまり，絶対数が正規化された大まかな出現傾向においては，同一の原曲歌手間における頻度分布は相対的に類似度が高かっ

表 3　各歌唱テクニックの生起頻度（楽譜情報に基づく 48 歌唱におけるボーカルパートの時間長総和は 3167.44 s）
Table 3　The occurrence frequency of each technique. The total duration of vocal parts, which is calculated by the note information, is 3167.44 seconds.

歌唱テクニック	ラベル数	全時間長 [s]	平均継続長 [s]
ビブラート	717	448.57	0.63
しゃくり	528	118.40	0.24
こぶし	144	33.14	0.23
フォール	140	31.25	0.23
ヒーカップ	126	20.35	0.16
メリスマ	38	16.36	0.44
ウィスパーボイス	11	54.5	4.95
ファルセット	86	96.16	1.14
ブレシー	52	41.57	1.03
ボーカルフライ	82	21.73	0.28
タングトリル	1	0.36	0.23
シャウト	2	1.16	0.39
セリフ語り	4	13.71	1.98

*4　Python モジュール scikit-learn を用いて算出．

© 2023 Information Processing Society of Japan

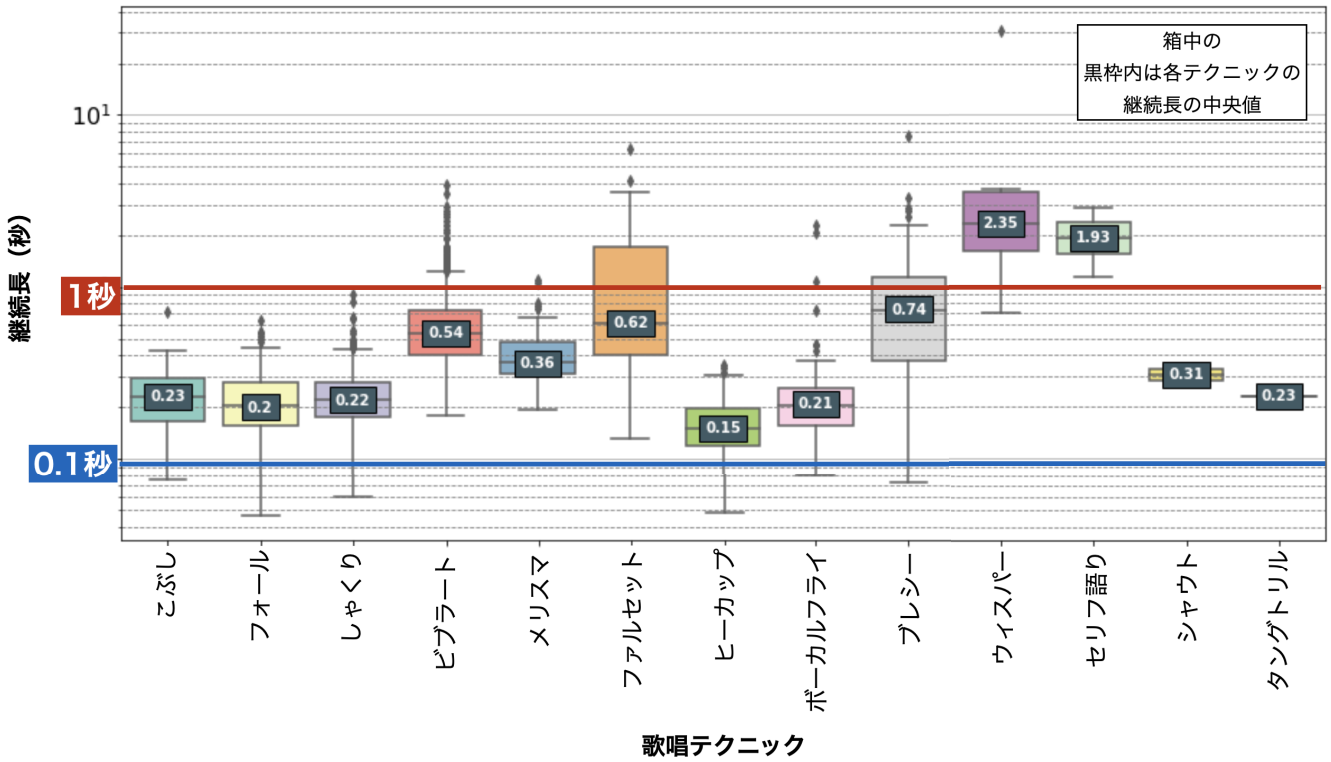

図 3　各歌唱テクニックの継続時間長の分布．縦軸は対数軸 [s]

Fig. 3　The distribution of duration of each singing technique. The vertical axis is seconds in log-scale.

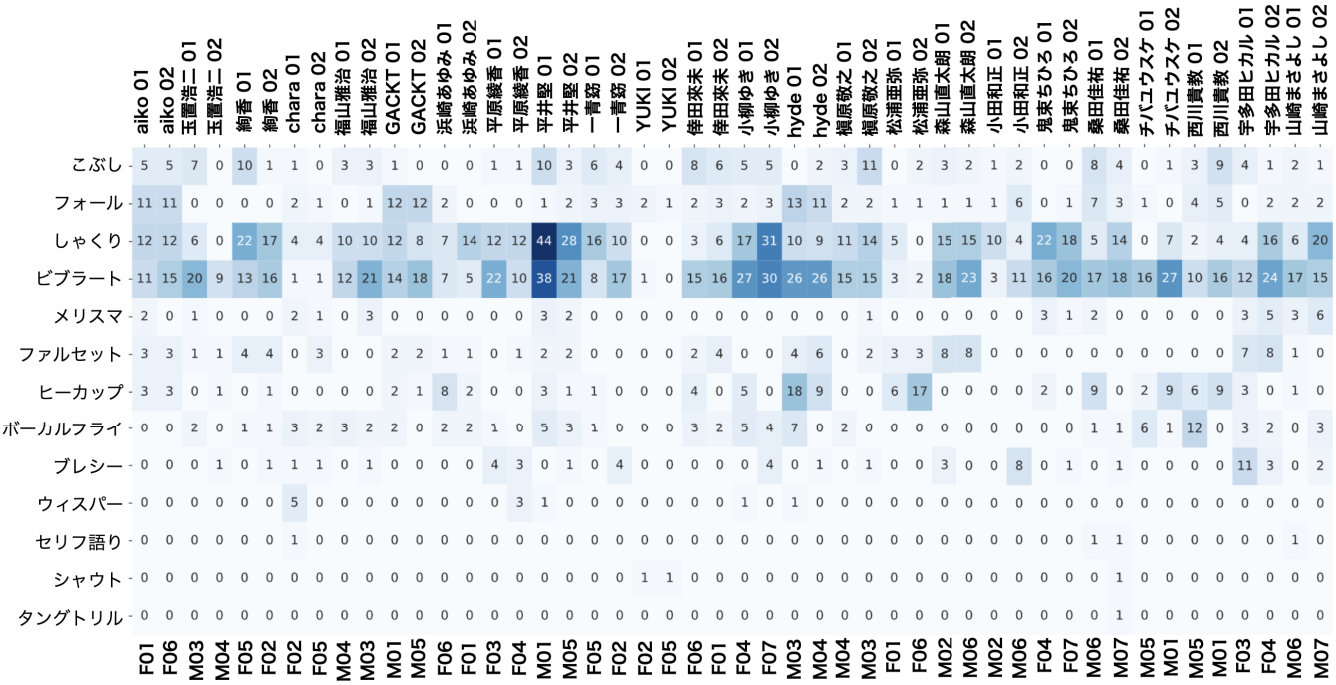

図 4　模倣歌手ごとの歌唱テクニックの生起頻度の可視化．模倣歌手名は「aiko 01」のように，原曲歌手名とその原曲歌手の何番目の模倣歌手かを示す模倣 ID で示し，1 人目の模倣歌手が aiko の「ボーイフレンド」を歌唱したことを意味する．また，下部の数字は，実際に模倣を実行した模倣歌手の ID を示す．頻度カウントはます内の数字のほか，ヒートマップ（色の濃淡）により表現している

Fig. 4　Frequency of singing techniques by imitation singers, sorted by original songs. The displayed singer names consist of original singer, and imitation singer number, respectively. For instance, “aiko 01” denotes that the original singer is “aiko” (singing “boy friend”), the first imitation singer of two. The value in each cell and tint of color indicates frequency counts.

© 2023 Information Processing Society of Japan

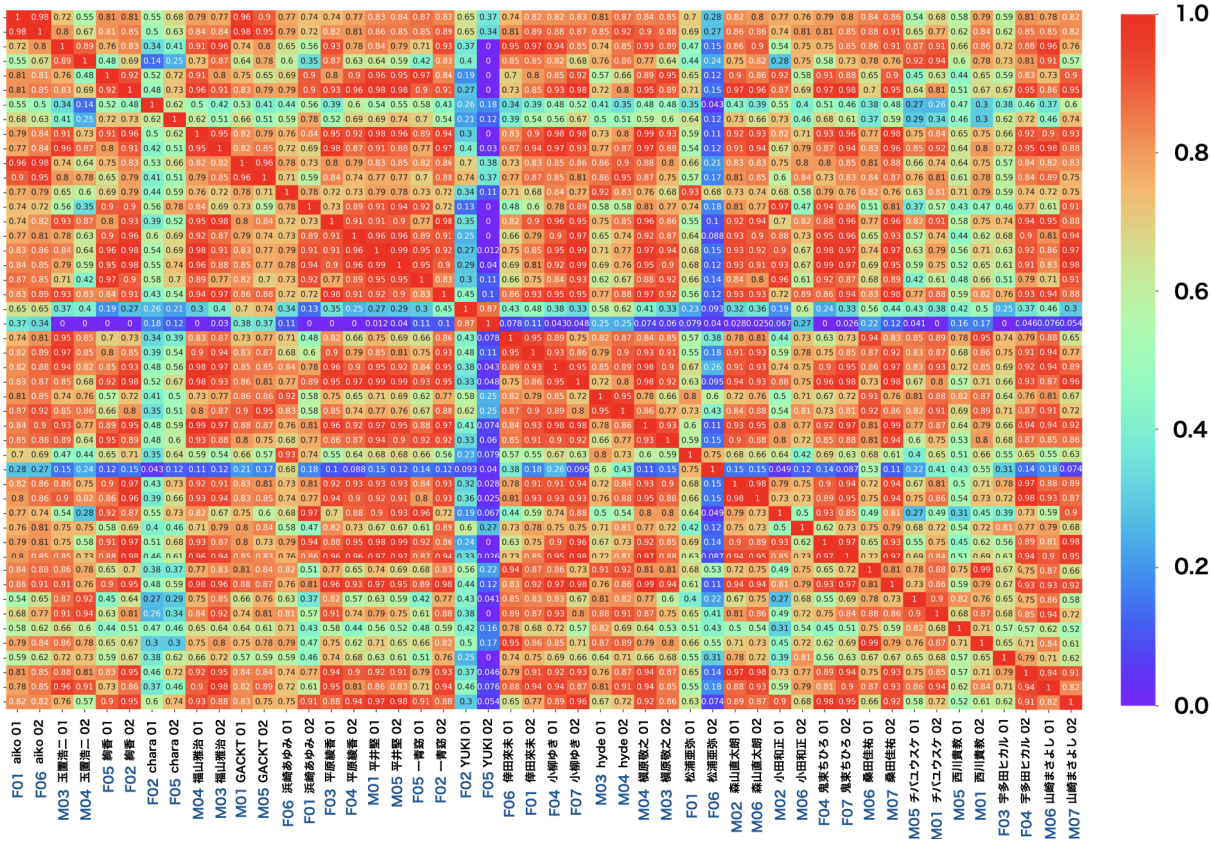

図 5　模倣歌唱におけるテクニック生起頻度のコサイン類似度．類似度の数値はます内の数字のほか，右のカラーバーに示す色相により表現している

Fig. 5　The cosine similarities of singing-technique occurrence frequencies in imitation singing. The value in each cell and hue indicates the similarity score.

たといえる．

また，類似する 2 人の原曲歌手の頻度分布間においてもコサイン類似度が高くなるという結果を得た．たとえば，平井堅と一青窈は，いずれの模倣歌手間においても 0.9 を超える結果となった．この両者は，±2 あるいは ±3 半音の音高シフトを行うことにより，声質が類似することが知られており，音響特徴量の観点からもそれが部分的に検証されているが [51]，歌唱テクニックの頻度分布間のコサイン類似度が高くなることを確認した．しかし，これらの結果は，楽曲の要素の影響を排除できておらず，今後各原曲歌手の楽曲数を増やし再検証することが必要である．

4.2　特徴（制御パラメータ）の分析

本節では，表 3 で最も多く生起したビブラートに着目し，そのパラメータとして深さ（extent）と速さ（rate）に関する分析結果を報告する．特に音高変動に焦点を当て，F0 を推定して分析する．F0 推定には CREPE [52] を用い，得られた値（f_{Hz}）から，単位を対数スケールである cent（f_{cent}）に変換して分析する．西洋平均律では，半音が 100 cent にあたる．中央ハ音の周波数（$f_c = 440 \times 2^{\frac{3}{12}-1} = 261.62\,\mathrm{Hz}$）の cent 値を 4800 cent とすると以下のように計算できる．

$$f_{\mathrm{cent}} = 1200 \log_2 \left(\frac{f_{\mathrm{Hz}}}{f_c} \right) + 4800 \tag{1}$$

ここで，F0 系列の時間分解能は 10 ms とした．

4.2.1　ビブラート

中野ら [16] の方法を参考に，以下のように各ビブラートパラメータを計算した（**図 6**）．まず，ビブラートのラベルが付与された区間の F0 系列について，極大値および極小値を計算することでピークを検出する．この区間内で検出できたピークの個数を N として，ビブラートパラメータの extent，rate は以下の式により計算する．

$$\mathrm{extent} = \frac{1}{N} \cdot \sum_{n=1}^{N} E_n, \qquad \frac{1}{\mathrm{rate}} = \frac{1}{N} \cdot \sum_{n=1}^{N} R_n \tag{2}$$

E_n はビブラート概形を考慮した変動幅，R_n は極大（極小）値間の時間差である．

全 48 歌唱における extent の平均は 181.3 cent，rate の平均は 6.53 Hz であった．また，模倣歌手ごとの extent と rate の分布を**図 7** に示す．図からは，両パラメータとも，原曲歌手が異なるとその中央値が異なることが分かる．

extent における各模倣歌手の中央値は，GACKT の模倣歌手 1（M01）が 363.4 cent，模倣歌手 2（M05）が 329.8 cent，西川貴教の模倣歌手 1（M05）が 298.2 cent であり，深いビ

© 2023 Information Processing Society of Japan

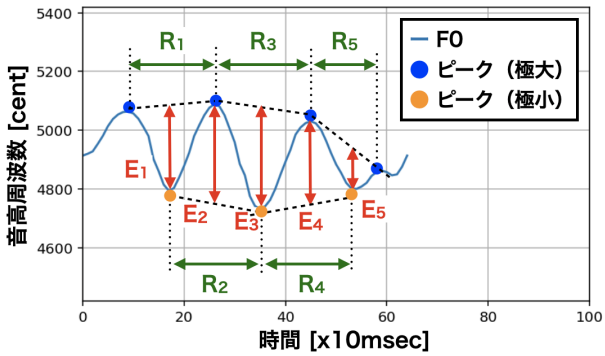

図 6　ビブラートの F0 系列と，ビブラートパラメータの extent（深さ）および rate（速さ）の計算のためのピーク（極大および極小）の系列

Fig. 6　An F0 sequence of vibrato and a sequence of peaks (local maxima and minima) to calculate vibrato parameters of extent (depth) and rate (speed).

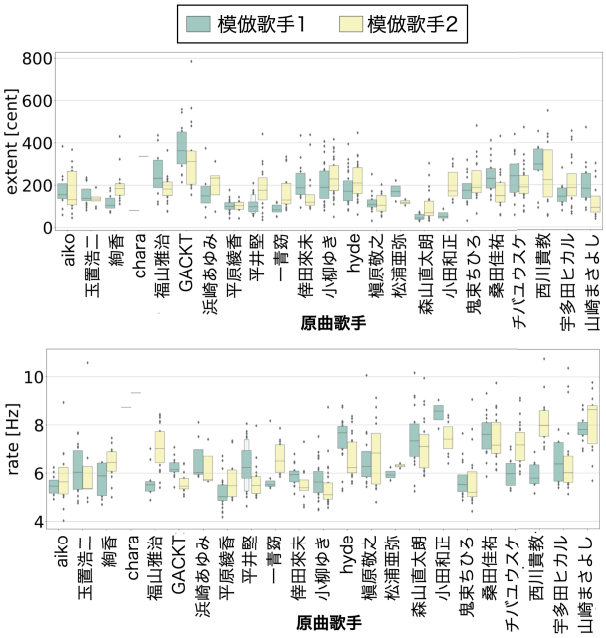

図 7　ビブラートパラメータ分布（注：原曲歌手 chara については，ビブラートが 1 カ所のみ検出されたため，E_n と R_n の中央値のみが可視化されている．原曲歌手 YUKI については図中にないが，模倣歌手 2 にはビブラートが存在せず，模倣歌手 1 については 1 カ所存在し，その extent は 301.0 cent, rate は 6.9 Hz であった

Fig. 7　The distribution of vibrato parameters. This shows only median for chara's imitation since the vibrato is observed once. As for YUKI's imitation, only the result of the imitation singer 1 (extent of 301.0 cent and rate of 6.9 Hz) is shown since the imitation singer 2 does not use vibrato.

ブラートが多く観測された．逆に，平原綾香の模倣歌手 1（F03）が 100.4 cent，模倣歌手 2（F04）が 104.5 cent，森山直太朗の模倣歌手 1（M02）が 43.7 cent であり，浅いビブラートが多く観測された．

rate においても，それぞれの中央値は，hyde の模倣歌手 1（M03）が 7.68 Hz，桑田佳祐の模倣歌手 1（M06）が 7.62 Hz，模倣歌手 2（M07）が 7.18 Hz，山崎まさよしの模倣歌手 2（M07）が 8.65 Hz と，速いビブラートを多用していたのに対し，aiko（模倣歌手 1（F01）：5.47 Hz，模倣歌手 2（F06）：5.65 Hz）や小柳ゆき（模倣歌手 1（F04）：5.64 Hz，模倣歌手 2（F07）：5.11 Hz），鬼束ちひろ（模倣歌手 1（F03）：5.53 Hz，模倣歌手 2（F04）：5.21 Hz）の模倣歌手は 6 Hz を超える速さのビブラートをほとんど用いていないことが分かった．

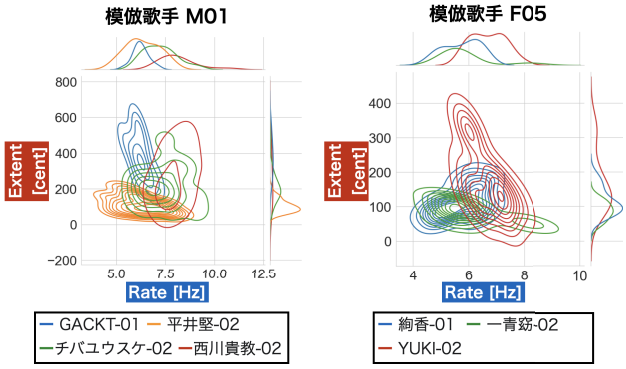

図 8　ビブラートパラメータ extent と rate の 2 次元分布（左：模倣歌手 M01，右：模倣歌手 F05）

Fig. 8　The plot of vibrato parameters (extent and rate) by imitation singers (left: singer-M01, right: singer-F05).

また，同一歌手が異なる原曲を歌い分けているかどうかを分析するため，男女の歌手各 1 名ずつについて，extent と rate の 2 次元分布をカーネル密度推定（KDE）を用いて可視化した結果を図 8 に示す．ここで，KDE プロットには，Python の可視化モジュールである seaborn を用いた．

これらの分析結果から，たとえば，原曲歌手を GACKT とする，2 名の模倣歌手（M01 と M05）の extent がどちらも中央値が高く幅広い（図 7）こと，またその一方で，別の原曲歌手の歌い方を模倣した場合にはパラメータの分布が異なる（図 8 右の M01）ことから，ビブラートパラメータが使い分けられている可能性が示唆される．Saitou ら [53] および齋藤ら [3] により，アマチュア歌唱者にとっては rate を変更しにくいことが報告されているが，プロ歌唱者においてはその傾向が異なる可能性がある．ただし，今回のデータには原曲歌手 1 名につき，1 曲のみを分析に用いたため，楽曲による影響が除けていないことに注意する必要がある．

今後は，同じ原曲歌手の歌い方を模倣した模倣歌手のビブラートパラメータが似ているのか，また，ある歌唱者が別の歌い方を模倣した場合にはビブラートパラメータ分布が異なるのか等について定量的な分析を進めたい．その際には，聴取実験による個々の歌唱に対する模倣の度合いの高さの評価や，原曲歌手に関する分析，そして原曲歌手の異なる楽曲に対する分析等が必要になると考えられる．

© 2023 Information Processing Society of Japan

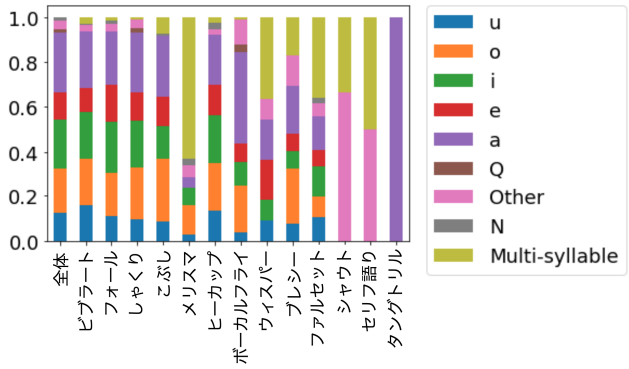

図 9　各歌唱テクニックにおいて，それが用いられる音素の頻度分布（注：「全体」は全音符の音素数．それ以外は，歌唱テクニックごとに頻度を正規化したため，音素数の分布に影響を受ける）

Fig. 9　The distribution of vowels on each singing technique. “All” denotes the cumulation of all phonemes of all notes. The occurrence distribution of each singing technique is affected by the distribution of phonemes since it is normalized.

4.3　楽譜情報を用いた生起箇所の分析

前節までで，全 13 種類の歌唱テクニックの頻度の分布を示し，ビブラートに関しては，さらにそのパラメータの分布を示した．本節では，楽曲のメロディの楽譜情報を用い，音楽要素である歌詞の音素（母音）の種類，音高，音高差，音長，フレーズ位置と歌唱テクニックの生起箇所の関係を分析する

4.3.1　歌詞の音素との関係

各歌唱テクニックについて，それが使われる際の歌詞の音素の分布を**図 9** に示す．主に母音の音素を集計対象とし，/a，e，i，o，u，Q（促音）[*5]，N（撥音），その他（英語等），複数音素/とした．ここで複数音素とは，音素の種類が歌唱テクニックラベルの区間内で複数存在する場合を意味する．また図には，楽譜情報を用いて全音符の母音の数を集計して正規化した分布もあわせて示した．歌唱テクニックラベルが音響信号に対して付与されたため，ラベルの時刻情報と楽譜の音符の発音時刻とを対応付ける必要がある．具体的には，歌唱テクニックラベルの区間内に存在する音符を割り当てるが，ラベル区間の一部分に音符が存在した場合は音符長の半分以上にラベルが存在するかどうかで決定した．

図 9 から，ビブラート，フォール，しゃくり，こぶし等の音高を制御する歌唱テクニックは，単一音素で生起する確率が高いことが分かる．

各テクニックの音素分布は，すべての音符の音素の頻度分布に類似した傾向を持っている場合があるが，今回の集計方法では，もともとの音素数から影響を受けるためである（歌詞に/a/が多ければ，それぞれのテクニックでも/a/が生起しやすい）．また，テクニック間で分布が異なるのは，音符の高さや長さ等に影響を受けている可能性もある．音素の頻度に影響されない分布の算出は，今後の課題である．また，メリスマは複数音素かつ音高変化をともなうことから 1 つのラベルに複数音素が含まれる結果となり，ウィスパー，ファルセット，ブレシー，シャウト等の声質制御に基づくテクニックも，複数音素にまたがることが多かった．

*5　促音部分の歌唱に特徴的な歌い方が生起する可能性を考え，「あっ」，「こっ」のように，促音が後続する場合については，促音としてもカウントして，分析対象に加えた．

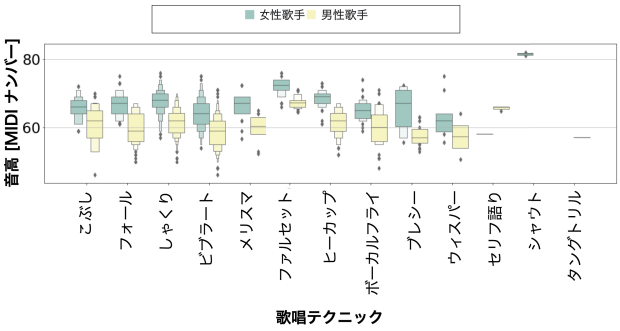

図 10　歌唱テクニックごとの音高の頻度分布

Fig. 10　The frequency distribution of pitch (MIDI note number) for each singing technique.

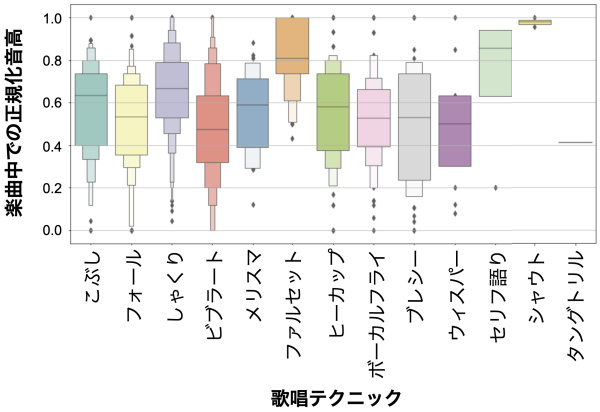

図 11　歌唱テクニックごとの音高の頻度分布（最低音と最高音を用いて正規化）

Fig. 11　The frequency distribution of pitch (MIDI note number) for each singing technique, normalized by the highest and lowest notes.

4.3.2　音符の音高との関係

各歌唱テクニックと音高との関係を分析するために，テクニックと対応付けられたラベルごとに楽譜の MIDI ノートナンバーを集計した分布を**図 10** に示す．図 10 では，声域の違いを正しく評価するために原曲歌手の性別で分けて可視化したが，楽曲の音域やキー変更の影響も除くことを意図して，歌唱者ごとの音高を正規化して集計した分布を**図 11** に示す．ここで，各音符における正規化音高 $\bar{x}$ は，元の音高を x，その楽曲の最高音と最低音の値をそれぞれ $x_{\max}$，$x_{\min}$ としたとき，以下のように求めた．

$$\bar{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{3}$$

© 2023 Information Processing Society of Japan

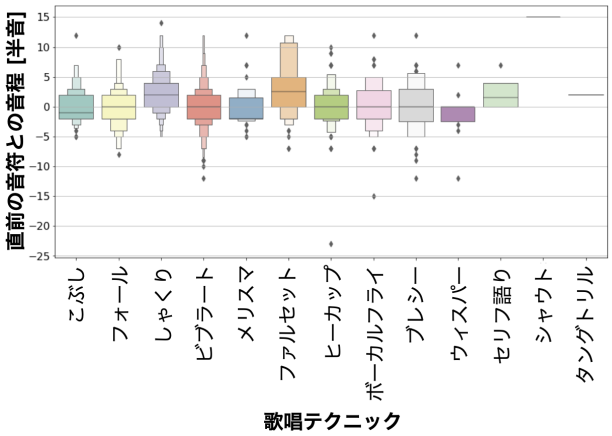

図 12　歌唱テクニックごとの直前の音との音程の頻度分布

Fig. 12　The frequency distribution of intervals with the preceding note for each singing technique.

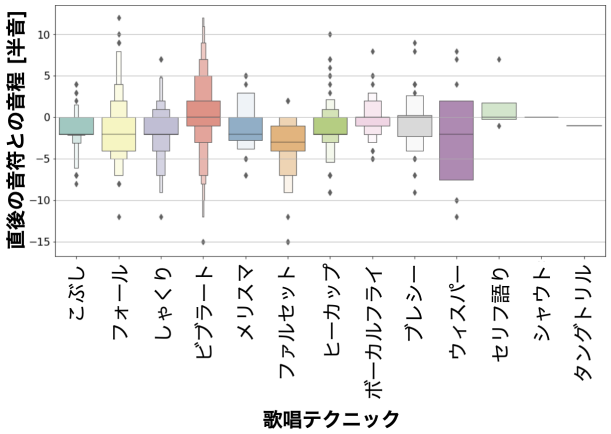

図 13　歌唱テクニックごとの直後の音との音程の頻度分布

Fig. 13　The frequency distribution of intervals with the following note for each singing technique.

これらの分布からは，ファルセットやシャウトが高い音高に割り振られることが分かる（図 10）．その他，しゃくりの音高分布の平均が比較的高かった．ラベル付けの際に，高い音へ歌い上げる際にこのテクニックを使っている主観的な印象を受けたが，今後，より詳細な分析が必要である．

4.3.3　前後の音高差との関係

前項と関連し，音高に関する分析のために前後の音符との音高差をそれぞれ**図 12**，**図 13** に示す．

ここでは，ファルセットとしゃくりにおいて前後の音高よりも高くなること，フォールが次の音で低くなることが分かる．これはテクニックの性質を正しく表していることから，フォールは次の低い音に滑らかに移行する目的で用いることが多いという性質が示唆される．ラベルと音符の対応付けが正しくなされていることも示唆している．また，ビブラートは分布が広いが，平均的には前後の音符に音高差がない場合に生成されることが多い．

4.3.4　音長との関係

歌唱テクニックと音符の時間長の関係性分析のために音符の音長に関する分布を**図 14** に示す．メリスマとウィスパーボイス，セリフ語りが長く，次にしゃくり，ビブラート，ファルセットが長い音符において生起していた．図 3 に示した実際の長さと関係することが分かる．また，こぶし，フォール，ヒーカップ，ボーカルフライは短い音符において多く生起していた．これらのテクニックは継続長が短く，短い音符においても活用可能であり，また実際に今回の楽曲でもそのような箇所に用いられたことが示唆される．

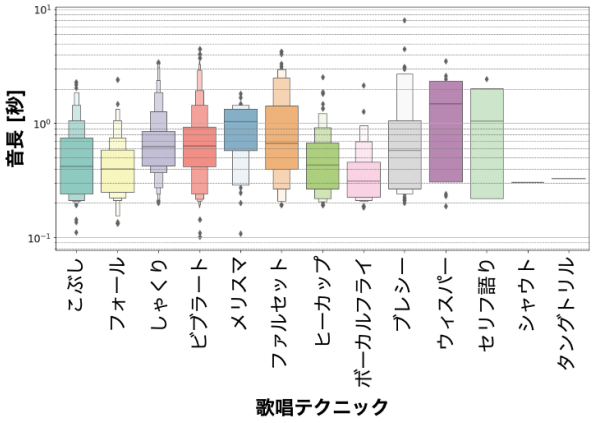

図 14　音符の時間長ごとの頻度分布

Fig. 14　The frequency distribution of note length for each singing technique.

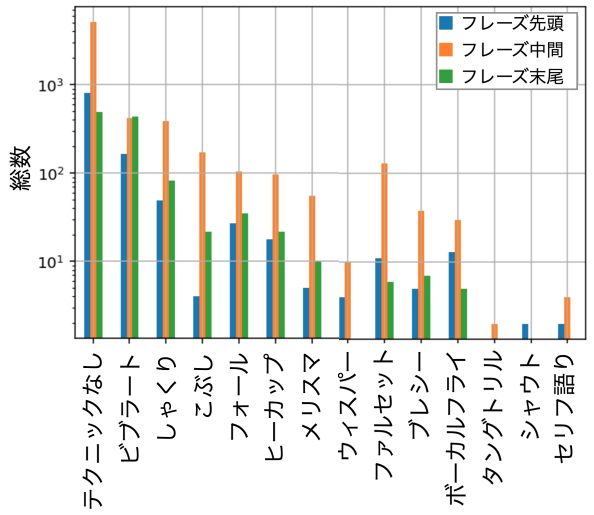

図 15　フレーズの最初と末尾の音符，それ以外（中間）の音符における歌唱テクニックの頻度分布．縦軸は対数軸

Fig. 15　The frequency distribution of singing techniques at musical notes of phrase positions (head, tail, and middle). The vertical axis is in the logarithmic scale.

4.3.5　音符のフレーズ内の位置との関係

各テクニックがフレーズのどの位置に生起しやすいかを分析するために，フレーズの先頭または末尾の音符で生起するか，それ以外の音符で生起するかを集計した．その分布を**図 15** に示す．本論文では，3.2 節で前述のとおり，16 分休符より長い休符を境に，個々のフレーズに自動分割した．個々のフレーズに自動分割した．歌唱テクニックラベ

© 2023 Information Processing Society of Japan

表 4　ビブラートパラメータと音高（基本周波数），ビブラート長との相関分析（**は有意水準 0.01 で相関があることを示す）

Table 4　Correlation analysis between pitch height (F0), vibrato duration, and vibrato parameters (rate and extent) (** denotes a correlation at a significance level of 0.01).

	音高（女性歌手）	音高（男性歌手）	正規化音高（全歌手）	ビブラート長（ラベル長）
extent	−0.424**	−0.336**	−0.31**	−0.127**
rate	0.225**	0.169**	0.025	−0.268**

ルは音響信号に基づいてラベル付けされていることから，複数の音符にまたがる可能性があるため，3 種類の分布の総数は実際の音符数とは異なる．

図からは，ビブラートはフレーズ途中もしくは終端で生起することが多く，ボーカルフライはフレーズの途中もしくは先頭で生起することが多いことが分かった．ビブラートはロングトーンの場合に，ボーカルフライ（エッジボイスとも呼ばれる）は音のアタック時に用いられることから，このような結果になった可能性がある．また，こぶしについは，フレーズの先頭で生起することが少なかった．

4.4　ビブラートパラメータとの関係

ビブラートパラメータについて，音高およびビブラート長（ラベル長）との関係を明らかにするため，相関分析を行った．パラメータ extent と rate について，音高（女性歌手），音高（男性歌手），正規化音高（男女），ビブラート長の 4 項目のピアソンの積率相関係数を**表 4** に示す．extent と全項目との間に負の相関が見られたことから，ビブラートが長いほど，音高が高いほど extent が浅くなる傾向にあることが分かる．また rate に関しては，音高（男女のいずれも）と正の相関が，ビブラート長と負の相関が見られた．つまり，音高が高いほど rate が速くなり，ビブラートが長いほど rate が遅くなる傾向がある．しかし，正規化音高と rate については有意な相関が見られなかった．これは，高い音高では rate が速くなるが，同一歌手が音高の違いで rate を変えるわけではない，ということを示唆している．ビブラート長と rate の相関には，楽曲のテンポが関係している可能性があるが，今後の検討課題である．

次に，歌手別の傾向について考察する．平井堅，小柳ゆきの各模倣歌手での事例について，**図 16** に extent-音高の例，**図 17** に extent-ビブラート長，**図 18** に rate-ビブラート長の例を示す．各歌手においても，extent-音高と，rate-ビブラート長における負の相関を確認できる．他の歌手でも似たような傾向を確認した．しかし，extent-ビブラート長においては，たとえば図 17 左からは，平井堅の模倣歌手 2 名については逆に弱い正の相関を確認できる．全体の相関値（−0.127）からも，必ずしも全歌手で確認できる事例ではなかったといえる．また，**図 19** に rate-音高の例を示す．小柳ゆきの模倣歌手 1 には正の相関が確認できるものの，その他の模倣歌手においては顕著な相関はないことが確認できる．他の歌手の例を確認したが，歌手に

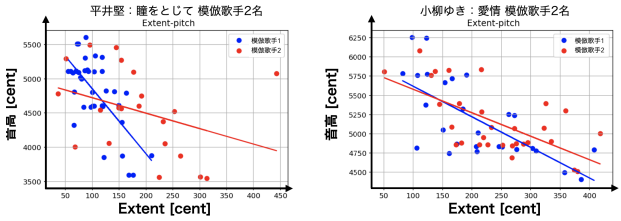

図 16　extent-音高の散布図およびその回帰直線（左：原曲歌手が平井堅，右：原曲歌手が小柳ゆき）

Fig. 16　Scatter plot of pitch and extent with its regression line (left: Ken Hirai's imitations, right: Yuki Koyanagi's imitations).

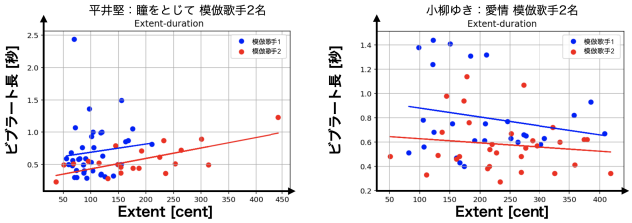

図 17　extent-ビブラート長の散布図およびその回帰直線（左：原曲歌手が平井堅，右：原曲歌手が小柳ゆき）

Fig. 17　Scatter plot of vibrato duration and extent with its regression line (left: Ken Hirai's imitations, right: Yuki Koyanagi's imitations).

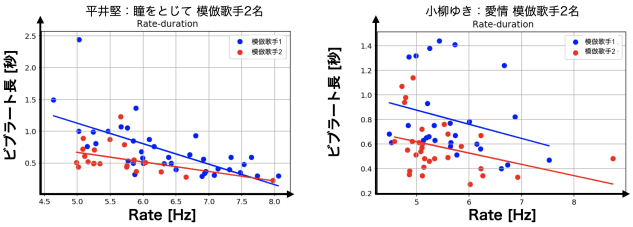

図 18　rate-ビブラート長の散布図およびその回帰直線（左：原曲歌手が平井堅，右：原曲歌手が小柳ゆき）

Fig. 18　Scatter plot of vibrato duration and rate with its regression line (left: Ken Hirai's imitations, right: Yuki Koyanagi's imitations).

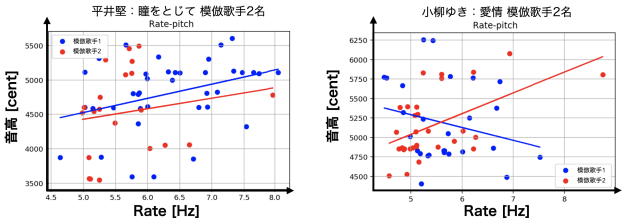

図 19　rate-音高の散布図およびその回帰直線（左：原曲歌手が平井堅，右：原曲歌手が小柳ゆき）

Fig. 19　Scatter plot of pitch and rate with its regression line (left: Ken Hirai's imitations, right: Yuki Koyanagi's imitations).

© 2023 Information Processing Society of Japan

よってこの相関は異なり，全歌手における共通な傾向は確認できなかった．

5. おわりに

本論文では J-POP において既存の歌手の歌い方を，別のプロ歌手が学術目的で模倣した歌声を用いて，歌唱テクニックの頻度と特徴，生起箇所について分析を行った．テクニックを複数種類分析し，楽譜情報と対応付けたことで，その特性を部分的に明らかにすることができた．また，歌い方の模倣に関しては，テクニックの頻度分布やビブラートパラメータの分布を通じてその妥当性の分析を行った．今後は，個々の分析の精度を高めるために，楽曲由来の影響の除去（たとえば，模倣歌手の人数や楽曲数の増加，異なる複数の模倣スタイルで同一の楽曲を歌唱したデータを用いる等）や，歌唱テクニックのアノテーション結果と楽譜との対応付けの高精度化，セクション（A メロ，サビ等）等の大局的な楽譜構造のアノテーション等が課題である．また，本研究では分析対象としていない原曲（伴奏付き歌唱）に関しては，バックコーラスやリバーブ等を含まない単旋律の歌声を精度高く音源分離可能な技術が将来登場すれば，その分析が可能になることが期待できる．それによって可能になる原曲歌唱と模倣歌唱間の比較も興味深い課題である．

謝辞 本論文の一部は JST 次世代研究者挑戦的研究プログラム（JPMJSP2124）の助成を受けて行われたものです．筑波大学大学院図書館情報メディア系の平賀譲名誉教授には，本論文に関して有益な助言をいただきました．

参考文献

[1] Demetriou, A., Jansson, A., Kumar, A. and Bittner, R.M.: Vocals in Music Matter: The Relevance of Vocals in the Minds of Listeners, *Proc. 19th International Society for Music Information Retrieval Conference* (*ISMIR 2018*), pp.514–520 (2018).

[2] Umbert, M., Bonada, J., Goto, M., Nakano, T. and Sundberg, J.: Expression Control in Singing Voice Synthesis: Features, approaches, evaluation, and challenges, *IEEE Signal Processing Magazine*, Vol.32, No.6, pp.55–73 (2015).

[3] 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，pp.571–574 (2011).

[4] 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，Vol.64, No.10, pp.616–623 (2008).

[5] Gupta, C., Li, H. and Goto, M.: Deep Learning Approaches in Topics of Singing Information Processing, *IEEE/ACM Trans. Audio, Speech, and Language Processing*, Vol.30, pp.2422–2451 (2022).

[6] Piao, Z. and Xia, G.: Sensing the Breath: A Multimodal Singing Tutoring Interface with Breath Guidance, *Proc. International Conference on New Interfaces for Musical Expression* (*NIME 2022*), pp.1–18 (2022).

[7] Humphrey, E.J., Reddy, S., Seetharaman, P., Kumar, A., Bittner, R.M., Demetriou, A., Gulati, S., Jansson, A., Jehan, T., Lehner, B., Kruspe, A. and Yang, L.: An Introduction to Signal Processing for Singing-voice Analysis: High Notes in the Effort to Automate the Understanding of Vocals in Music, *IEEE Signal Processing Magazine*, Vol.36, No.1, pp.82–94 (2018).

[8] 才野慶二郎：歌声の合成における応用技術—歌声合成システム，日本音響学会誌，Vol.75, No.7, pp.406–411 (2019).

[9] 中野倫靖，後藤真孝：歌声の合成における基盤技術—歌声合成における特徴量の制御，日本音響学会誌，Vol.75, No.7, pp.400–405 (2019).

[10] Cho, Y.-P., Yang, F.-R., Chang, Y.-C., Cheng, C.-T., Wang, X.-H. and Liu, Y.-W.: A Survey on Recent Deep Learning-driven Singing Voice Synthesis Systems, *2021 International Conference on Artificial Intelligence and Virtual Reality* (*AIVR*), pp.319–323, IEEE (2021).

[11] Sundberg, J.: *The Science of the Singing Voice*, Northern Illinois University Press (1987).

[12] Seachore, C.E.: A Musical Ornament, the Vibrato, *Psychology of Music*, McGraw-Hill (1938).

[13] Prame, E.: Measurements of the Vibrato Rate of Ten Singers, *The Journal of the Acoustical Society of America*, Vol.96, No.4, pp.1979–1984 (1994).

[14] Prame, E.: Vibrato Extent and Intonation in Professional Western Lyric Singing, *The Journal of the Acoustical Society of America*, Vol.102, pp.616–621 (1997).

[15] Saitou, T., Unoki, M. and Akagi, M.: Development of an F0 Control Model Based on F0 Dynamic Characteristics for Singing-Voice Synthesis, *Speech Communication*, Vol.46, No.3-4, pp.405–417 (2005).

[16] 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).

[17] 鈴木千文，坂野秀樹，旭　健作，森勢将雅：歌唱音声の類似度評価を目的とした基本周波数の変動量を反映するビブラート特徴量の提案，電気学会論文誌 C，Vol.137, No.12, pp.1607–1614 (2017).

[18] Ikemiya, Y., Itoyama, K. and Okuno, H.G.: Transferring Vocal Expression of F0 Contour Using Singing Voice Synthesizer, *Proc. 27th International Conference on Industrial, Engineering Other Applications of Applied Intelligent Systems* (*IEA/AIE 2014*), pp.250–259 (2014).

[19] Ardaillon, L., Degottex, G. and Roebel, A.: A Multi-Layer F0 Model for Singing Voice Synthesis Using a B-Spline Representation with Intuitive Controls, *Proc. 16th Annual Conference of International Speech Communication Association* (*INTERSPEECH 2015*), pp.3375–3379 (2015).

[20] Yang, L.: Computational Modelling and Analysis of Vibrato and Portamento in Expressive Music Performance, PhD Thesis, Queen Mary University of London (2017).

[21] Hakes, J., Shipp, T. and Doherty, E.T.: Acoustic Characteristics of Vocal Oscillations: Vibrato, Exaggerated Vibrato, Trill, and Trillo, *Journal of Voice*, Vol.1, No.4, pp.326–331 (1988).

[22] Gupta, C., Li, H. and Wang, Y.: Perceptual Evaluation of Singing Quality, *Proc. 9th Asia Pacific Signal and Information Processing Association Annual Summit and Conference* (*APSIPA ASC 2017*), pp.577–586 (2017).

[23] Gupta, C., Li, H. and Wang, Y.: A Technical Framework for Automatic Perceptual Evaluation of Singing Quality, *APSIPA Trans. Signal and Information Processing*, Vol.7, pp.1–11 (2018).

[24] 右田尚人，森勢将雅，西浦敬信：歌唱データベースを用いたヴィブラートの個人性の制御に有効な特徴量の検討，

© 2023 Information Processing Society of Japan

情報処理学会論文誌，Vol.52, No.5, pp.1910–1922 (2011).

[25] Miyazaki, T. and Morise, M.: Building a Measurement Model for Simulating Naturalness of Vibrato based on Subjective Evaluation, *IEICE Trans. Information and Systems*, Vol.E104-D, No.4, pp.521–525 (2021).

[26] Lee, Y., Oya, M., Kaburagi, T., Hidaka, S. and Nakagawa, T.: Differences Among Mixed, Chest, and Falsetto Registers: A Multiparametric Study, *Journal of Voice* (2021).

[27] 平山健太郎，伊藤克亘：ポピュラー歌唱における高音域の声区と発声状態の判別手法，情報処理学会研究報告音声言語情報処理（SLP），Vol.2012, No.16, pp.1–6 (2012).

[28] 小島　俊，齋藤　毅，中野倫靖，後藤真孝，三好正人：歌声における裏声と地声を識別するための音響特徴量の検討，電子情報通信学会技術研究報告，EA, 応用音響，Vol.112, No.266, pp.67–72 (2012).

[29] Stoller, D. and Dixon, S.: Analysis and Classification of Phonation Modes in Singing, *Proc. 17th International Society for Music Information Retrieval Conference* (*ISMIR 2016*), pp.80–86 (2016).

[30] 山本雄也：歌唱テクニックの自動検出に向けた技術基盤の構築，筑波大学（情報学）修士論文 (2021).

[31] Wilkins, J., Seetharaman, P., Wahl, A. and Pardo, B.A.: VocalSet: A Singing Voice Dataset, *Proc. 19th International Society for Music Information Retrieval Conference* (*ISMIR 2018*), pp.468–474 (2018).

[32] 山本雄也，Nam, J.，寺澤洋子，平賀　譲：歌唱テクニックの識別における hand-cratfted 特徴量と深層学習抽出特徴量の比較，情報処理学会研究報告音楽情報科学（MUS），Vol.2021-MUS-130, No.30, pp.1–5 (2021).

[33] Kim, K.L., Lee, J., Kum, S., Park, C.L. and Nam, J.: Semantic Tagging of Singing Voices in Popular Music Recordings, *IEEE/ACM Trans. Audio, Speech, and Language Processing*, Vol.28, pp.1656–1668 (2020).

[34] Kalbag, V. and Lerch, A.: Scream Detection in Heavy Metal Music, *Proc. 19th Sound and Music Computing Conference* (*SMC 2022*), pp.403–410 (2022).

[35] Yamamoto, Y., Nam, J. and Terasawa, H.: Analysis and Detection of Singing Techniques in Repertoires of J-POP Solo Singers, *Proc. 23rd International Society for Music Information Retrieval Conference* (*ISMIR 2022*), pp.384–391 (2022).

[36] Ohishi, Y., Kameoka, H., Mochihashi, D. and Kashino, K.: A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components, *Proc. 13th Annual Conference of International Speech Communication Association* (*INTERSPEECH 2012*), pp.474–477 (2012).

[37] 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. 10th International Society for Music Information Retrieval Conference* (*ISMIR 2009*), pp.393–397 (2009).

[38] Panteli, M., Bittner, R., Bello, J.P. and Dixon, S.: Towards the characterization of singing styles in world music, *Proc. 42nd IEEE International Conference on Acoustics, Speech, and Signal Processing* (*IEEE ICASSP 2017*), pp.636–640 (2017).

[39] Abeßer, J. and Müller, M.: Fundamental Frequency Contour Classification: A Comparison between Hand-crafted and CNN-based Features, *Proc. 44th IEEE International Conference on Acoustics, Speech, and Signal Processing* (*IEEE ICASSP 2019*), pp.486–490 (2019).

[40] Repetto, R.C., Gong, R., Kroher, N. and Serra, X.: Comparision of the Singing Style of Two Jingju Schools, *Proc. 16th International Society for Music Information Retrieval Conference* (*ISMIR 2015*), pp.507–513 (2015).

[41] Miryala, S.S., Bali, K., Bhagwan, R. and Choudhury, M.: Automatically Identifying Vocal Expressions for Music Transcription, *Proc. 14th International Society for Music Information Retrieval Conference* (*ISMIR 2013*), pp.239–244 (2013).

[42] Mauch, M., Frieler, K. and Dixon, S.: Intonation in Unaccompanied Singing: Accuracy, Drift and a Model of Reference Pitch Memory, *The Journal of the Acoustical Society of America*, Vol.136, pp.1–11 (2014).

[43] Nichols, E., Morris, D., Basu, S. and Raphael, C.: Relationships Between Lyrics and Melody in Popular Music, *Proc. 10th International Society for Music Information Retrieval Conference* (*ISMIR 2009*), pp.471–476 (2009).

[44] Arai, M., Matuba, T., Hashida, M. and Katayose, H.: Revealing the Secret of "Groove" Singing: Analysis of J-pop Music, *Proc. 13th Sound and Music Computing Conference* (*SMC 2016*), pp.21–26 (2016).

[45] 中里南子：J・ポップにみられる装飾的旋律の歌い方—平井堅・桑田佳祐・ケミストリー・ドリカムの「コブシ」の分析を通して，音楽教育実践ジャーナル，Vol.5, No.1, pp.32–39 (2007).

[46] 山田知彦, 武藤　聡, 南角吉彦, 酒向慎司, 徳田恵一：HMM に基づく歌声合成のためのビブラートモデル化，情報処理学会研究報告音楽情報科学（MUS），Vol.2009, No.5, pp.1–6 (2009).

[47] Bonada, J. and Blaauw, M.: Hybrid Neural-Parametric F0 Model for Singing Synthesis, *Proc. 45th IEEE International Conference on Acoustics, Speech, and Signal Processing* (*IEEE ICASSP 2020*), pp.7244–7248 (2020).

[48] Hono, Y., Murata, S., Nakamura, K., Hashimoto, K., Oura, K., Nankaku, Y. and Tokuda, K.: Recent Development of the DNN-based Singing Voice Synthesis System — Sinsy, *Proc. 10th Asia Pacific Signal and Information Processing Association Annual Summit and Conference* (*APSIPA ASC 2018*), pp.1003–1009 (2018).

[49] Hono, Y., Hashimoto, K., Oura, K., Nankaku, Y. and Tokuda, K.: Sinsy: A Deep Neural Network-based Singing Voice Synthesis System, *IEEE/ACM Trans. Audio, Speech, and Language Processing*, Vol.29, pp.2803–2815 (2021).

[50] Cannam, C., Landone, C. and Sandler, M.: Sonic visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files, *Proc. 18th ACM International Conference on Multimedia* (*ACMMM 2010*), pp.1467–1468 (2010).

[51] Nakano, T., Yoshii, K. and Goto, M.: Vocal Timbre Analysis Using Latent Dirichlet Allocation and Cross-Gender Vocal Timbre Similarity, *Proc. 39th IEEE International Conference on Acoustics, Speech, and Signal Processing* (*IEEE ICASSP 2014*), pp.5239–5243 (2014).

[52] Kim, J.W., Salamon, J., Li, P. and Bello, J.P.: CREPE: A Convolutional Representation for Pitch Estimation, *Proc. 43rd IEEE International Conference on Acoustics, Speech, and Signal Processing* (*IEEE ICASSP 2018*), pp.161–165 (2018).

[53] Saitou, T. and Goto, M.: Acoustic and Perceptual Effects of Vocal Training in Amateur Male Singing, *Proc. 10th Annual Conference of International Speech Communication Association* (*INTERSPEECH 2009*), pp.832–835 (2009).

© 2023 Information Processing Society of Japan

山本 雄也　（学生会員）

2019 年筑波大学情報学群情報メディア創成学類卒業．2021 年同大学大学院図書館情報メディア研究科博士前期課程修了．修士（情報学）．現在，同大学院人間総合科学学術院博士後期課程．

中野 倫靖　（正会員）

2008 年筑波大学大学院図書館情報メディア研究科博士後期課程修了．博士（情報学）．現在，産業技術総合研究所主任研究員．日本音響学会会員．2009 年情報処理学会山下記念研究賞（音楽情報科学研究会），2013 年 Sound and Music Computing Conference（SMC 2013）The Best Paper Award，2018 年 IEEE Pacific Visualization Symposium（PacificVis 2018）Honorable Mention Poster Award 等各受賞．

後藤 真孝　（正会員）

1998 年早稲田大学大学院理工学研究科博士後期課程修了．博士（工学）．現在，産業技術総合研究所首席研究員．2009〜2017 年に IPA 未踏 IT 人材発掘・育成事業 PM，2016〜2022 年に JST ACT-I「情報と未来」研究総括を兼任．現在，日本学術会議連携会員，統計数理研究所客員教授，筑波大学大学院教授（連携大学院）等を兼任．日本学士院学術奨励賞，日本学術振興会賞，ドコモ・モバイル・サイエンス賞基礎科学部門優秀賞，市村学術賞，FIT 船井業績賞，科学技術分野の文部科学大臣表彰若手科学者賞，星雲賞等，64 件受賞．本会フェロー．

寺澤 洋子　（正会員）

電気通信大学電子工学科・同大学院修士課程電子工学専攻修了．スタンフォード大学音楽学科 CCRMA 修士課程および博士課程修了．Ph.D.（Music）．筑波大学図書館情報メディア系准教授．

© 2023 Information Processing Society of Japan