

# ポピュラー音楽における模倣歌唱を用いた 歌唱テクニックの頻度・特徴・生起箇所の分析

山本 雄也<sup>1,a)</sup> 中野 倫靖<sup>2,b)</sup> 後藤 真孝<sup>2,c)</sup> 寺澤 洋子<sup>1,d)</sup> 平賀 譲<sup>1,e)</sup>

**概要:** 本研究は、ポピュラー音楽における歌手の個性を構成する要素の一つである歌唱テクニックに着目し、その音響的特徴と楽譜上の出現傾向を明らかにすることを目的とする。従来、ビブラート等の一部の歌唱テクニックの特徴を分析する研究は行われてきたが、網羅的にかつ楽譜と対応付けた分析は行われていなかった。そこで本稿では、ポピュラー音楽のプロ歌手 24 名（男女各 12 名）の歌い方を対象として、楽曲のどの部分に、どの種類のテクニックを、どのように歌っているかを分析した結果を報告する。分析には、上記歌手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を用い、13 種類の歌唱テクニックに関する歌手毎の頻度分布とその分布間類似度、また歌手毎のビブラートパラメータ（extent と rate）の分布を示す。さらに、全テクニックの頻度とビブラートパラメータについて、それらと音高・音長（ビブラート長）との関係を分析し、歌唱テクニックがフレーズ上のどこで生じやすいかも分析した。

## 1. はじめに

本研究では、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、歌手の個性を構成する要素としての「歌唱テクニック」の分析を行う。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）は、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が知られていることから [2]、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、歌唱技能を向上させる支援技術、音楽情報検索技術、歌声合成技術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（例えば、[3-6]）、パラメータ化に関する検討やその応用が研究されてきた（例えば、[7-12]）。それに対して本研究では、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上で、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、可能な限り網羅的に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解明や、歌声情報処理システムへの応用へつなげる。

ポピュラー音楽の特性を反映するために、商業音楽において良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分析する。これにより、歌手の個性が歌唱テクニックとどう関係して知覚されるのか、また別の歌手によってそれがどう生成されるのか、という観点からの分析につながる可能性がある。したがってこのような方策は、人の歌声知覚・生成を理解する上でも重要な取組みである [2]。

## 2. 関連研究

本章では、歌唱テクニックの分析や生成に関連する研究や文献を分類してまとめ、本研究との差異を論ずる。

### 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（例えば、[3,5-7,13]）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

<sup>1</sup> 筑波大学

<sup>2</sup> 産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology (AIST)

a) s2130507@s.tsukuba.ac.jp

b) t.nakano@aist.go.jp

c) m.goto@aist.go.jp

d) terasawa@slis.tsukuba.ac.jp

e) hiraga@slis.tsukuba.ac.jp

に生じる揺らぎの回数)の二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 [8], Gupta *et al.* 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 [14,15]. Migita *et al.* 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 [16]. また Yang *et al.* 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なって京劇 (Beijing Opera) の歌唱スタイルの違いを分析した [12]. Ikemiya *et al.* は、ビブラート以外のテクニックとして、F0 軌跡の形状で特徴づけられるこぶし・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 [10].

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 *et al.* 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 など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 [17]. 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一つである。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 [18-20].

歌唱テクニックに関するデータセットも存在し、Wilkins *et al.* 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 [21]. 山本らは、VocalSet を用いて MFCC 及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 [22]. 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 [23].

## 2.2 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 *et al.* 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 [7]. また大石らは、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 [24].

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 *et al.* は基本周波数 (F0) とその微分

( $\Delta F0$ ,  $\Delta\Delta F0$ ) 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 [25]. また Ardaillon *et al.* 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 [11,26].

歌唱スタイルやジャンル、歌唱力や経験等を分析する上で、歌い方の違いが議論される事例もある。Panteli *et al.* 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 [27], AbeBer *et al.* 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 [28]. Caro *et al.* 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 [29]. また Miryala *et al.* 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 [30]. その他、Mauch *et al.* 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 [31].

## 2.3 歌唱と楽譜情報を関連付けた分析と応用

歌唱テクニックや歌い方に関して、楽譜における文脈を伴って分析した研究がある。Nichols *et al.* 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などの言語的特徴と、小節内位置や長さなどの音楽的特徴の関係を分析した [32]. また、Arai *et al.* 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などを示した [26]. その他、音響分析を伴わないが、中里らは 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 [33].

このように、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。例えば、歌声合成への応用が可能であり、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 [34]. また Wada *et al.* は、音楽的文脈を入力することで歌声の F0 軌跡生成を行った [35]. Hono *et al.* は歌声合成において、ビブラート・楽譜からのずれ・音素長さを明示的にモデル化した [36].

## 2.4 本研究の位置付け

以上から、本研究は網羅的にテクニックを扱い、かつ楽譜情報と対応付けて分析する点において、特長を持つ。

## 3. 対象とする歌唱データと歌唱テクニック

本研究では既存の歌唱を模倣した歌唱を用いるため、そ

表 1 AIST-SIDB の日本語歌手（キーの変更がある場合には、歌手名の横にその値を示す）

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

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 歌唱データ及び楽譜データ

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、ポピュラー音楽におけるプロ歌手の歌い方を、別のプロ歌手が模倣した歌声を学術目的で収録して構築したもので、表 1 に示す 24 名の歌手の歌い方を対象とした。表に示すように、歌手（楽曲）毎に 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、本稿では便宜上、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 の歌手を単に「歌手」と呼んで区別する。

AIST-SIDB の歌手は、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（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 60mm（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。

分析のための楽譜情報は、プロの音楽家 1 名が MusicXML の形式で原曲に対して採譜した。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\*1、移調せずに全て原音

\*1 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

表 2 本稿で扱う歌唱テクニック

歌唱テクニック名 (ラベル名)	変動する要素	変動の様子
ビブラート (vibrato)	音高・音量	揺らすように歌い、周期振動をつける
しゃくり (shakuri)	音高	音高を上連続変化させる
こぶし (kobushi)	音高	U 字/逆 U 字状に音高を連続変化させる
フォール (fall)	音高	音高を下連続変化させる
ヒーカップ (nakibushi)	音高・音色	しゃくりのように喉を締める
メリスマ (melisma)	音高	1 音節に複数の音高を割り当てる
ボーカルフライ (vocalfry)	声質	ざらついた音を出す
ファルセット (falsetto)	声質	裏声区で歌唱する
ブレイジー (breathy)	声質	呼息音を混入する
ウィスパーボイス (whisper)	声質	囁くように歌唱する
シャウト (shout)	-	叫ぶ
セリフ語り (spoken)	-	話すように歌う
タンゴトリル (tonguetrill)	-	巻き舌を入れる

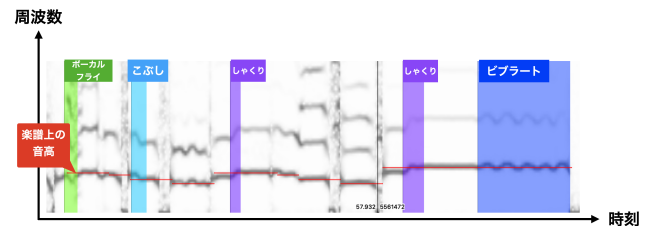


図 1 歌唱テクニックのラベル例（時刻及び種類）

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、全て連符で記した。しゃくりやフォールなど、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。

### 3.2 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に、第一著者が文献調査を行った。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）の観点から 32 種類が得られた [37]。続いて、前述した歌唱データに対してラベル付けを行い（方法は後述）、出現した歌唱テクニックを本稿における対象とする。そのようにして決定した歌唱テクニックを表 2 に示す。

### 3.3 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出現時刻をラベルとし、第一著者（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 [38] 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 1 にラベル例を示す。

## 4. 分析

各歌唱テクニックの歌手毎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

に、クリック（メトロノーム）を全く使用しないか、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。

表 3 各歌唱テクニックの出現頻度（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 3167.44s）

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
プレシー	52	41.57	1.03
ボーカルフライ	82	21.73	0.28
タンゴトリル	1	0.36	0.23
シャウト	2	1.16	0.39
セリフ語り	4	13.71	1.98

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 2 に示す。これらはポピュラー音楽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。多くのテクニックラベルの継続長は 0.1-1 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プレシー等の声質の制御によるテクニックや、セリフ語り等の楽曲依存のテクニックは、1 秒以上の長い継続長を持つ場合があった。

一方、歌手毎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 3 に、頻度分布間のコサイン類似度<sup>\*2</sup>を図 4 に示す。図 4 から、歌手 1 と歌手 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は tmr\_heatcapacity（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は oda\_kirakira（原曲歌手は小田和正）の 0.5 であった。図 3 より、前者についてはボーカルフライに差があり、後者についてはプレシー・しゃくり・ビブラートに差があることが分かる。したがって、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、同一の原曲歌手の歌手間における頻度分布が類似しているということが言える。

#### 4.2 特徴（制御パラメータ）の分析

本節では、表 3 で最も多く出現したビブラートに着目し、

\*2 Python モジュール scikit-learn を用いて算出。

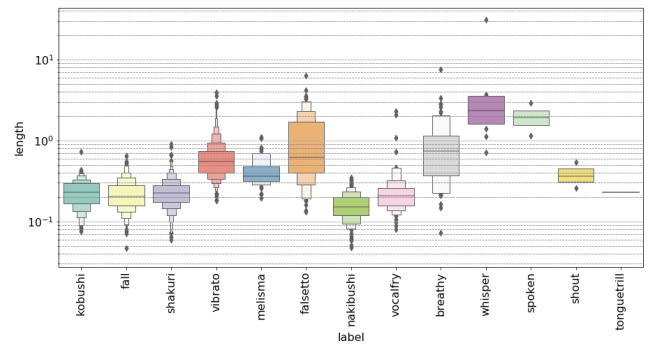


図 2 各歌唱テクニックの継続時間長の分布。縦軸は対数軸 [s]。

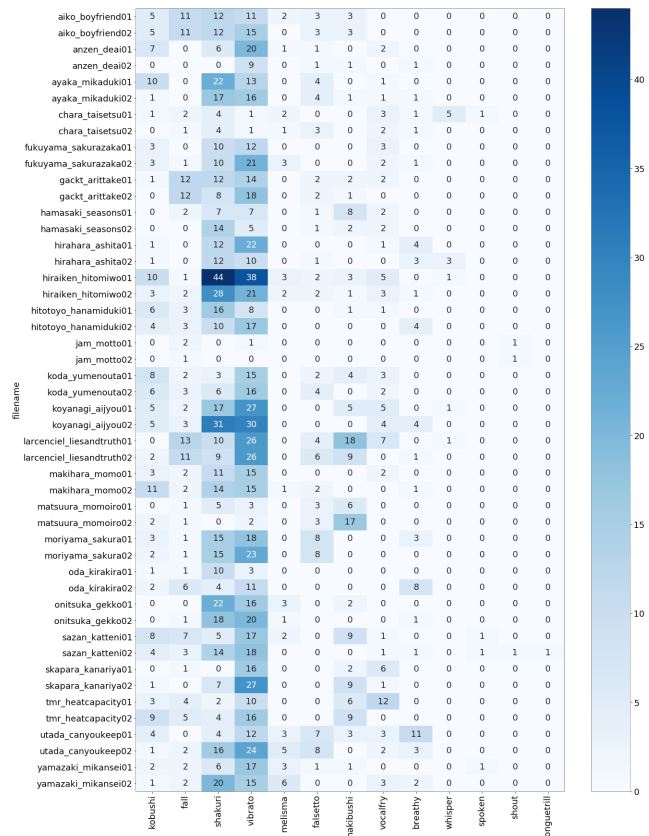


図 3 歌手ごとの歌唱テクニックの出現頻度の可視化。

その深さ (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 [39] を使い、得られた値 ( $f_{\text{Hz}}$ ) から、単位を対数スケールである cent ( $f_{\text{cent}}$ ) に変換して分析する。西洋平均律では、半音が 100cent にあたる。中央ハ音の周波数 ( $f_c = 440 \times 2^{\frac{3}{12}} = 261.62... \text{Hz}$ ) の cent 値を 4800cent とすると以下のように計算できる。

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

##### 4.2.1 ビブラート

ビブラートについては、深さと速さについて分析する。ビブラートラベルが付与された時刻の F0 からそのピーク (極大値と極小値) を検出し、中野ら [8] 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。

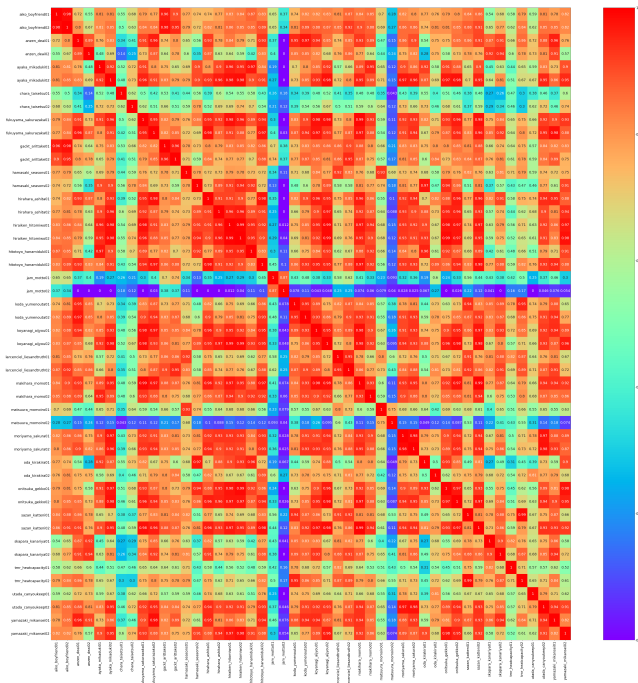


図 4 出現頻度のコサイン類似度.

$$\text{extent} = \frac{1}{N} \cdot \sum_{n=1}^N E_n, \quad \frac{1}{\text{rate}} = \frac{1}{N} \cdot \sum_{n=1}^N R_n \quad (2)$$

ここで  $N$  はピーク個数,  $E_n$  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 と rate の分布を図 5 に示す. 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 6 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, 例えば, 原曲歌手を Gackt とする 2 名の歌手 (M01 と M05) の extent がどちらも中央値が高く幅広い (図 5) 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 6 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 *et al.* [40] 及び齋藤ら [2] 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, プロ歌手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, ある歌手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などが必要になると考えられる.

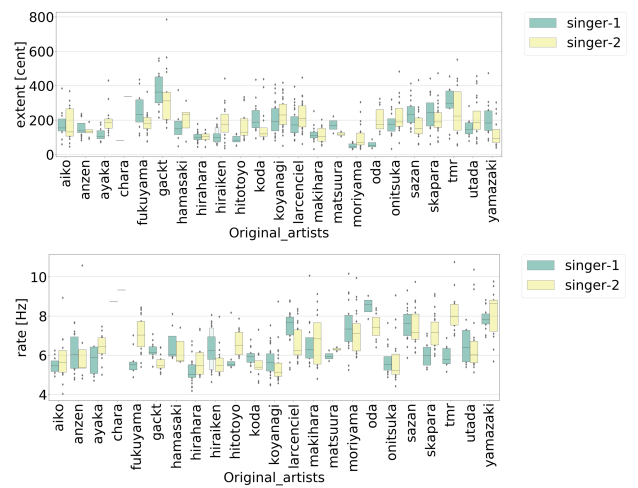


図 5 ビブラートパラメータ分布 (注: 原曲歌手 chara については,  $E_n$  と  $R_n$  が 1 か所のみ検出されたため, 中央値のみが可視化されている. 原曲歌手 jam については, 歌手 2 にビブラートが存在しなかったためプロットされなかったが, 歌手 1 については extent は 301.0cent, rate は 6.9Hz であった.)

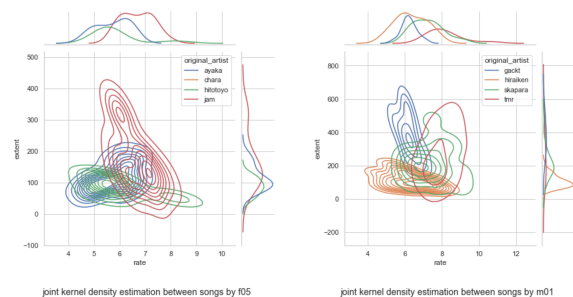


図 6 ビブラートパラメータ extent と rate の 2 次元分布 (左: 歌手 F05, 右: 歌手 M01)

### 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, それらを楽譜情報と対応付け, 歌詞の音素 (母音) の種類, 音高, 音長, フレーズ位置との関係を分析する.

#### 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 7 に示す. 音素は主に母音を対象とし, /a, e, i, o, u, Q (促音), 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 (All) も併せて示した. 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情報とを対応付ける必要がある. 具体的には, 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. ただ

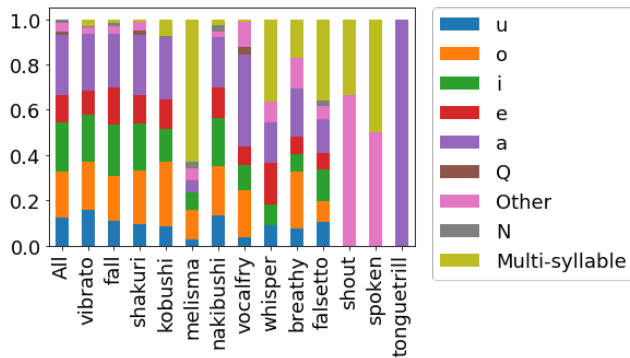


図 7 各歌唱テクニックにおいて、それが用いられる音素の頻度分布（注：「All」は全音素の音素数。それ以外は、歌唱テクニック毎に頻度を正規化したため、音素数の分布に影響を受ける。）

し、上記のみでは対応できないテクニックには例外処理を設けた。例えば、しゃくりは本来の音符よりも早く生起することが多いため、ラベル時刻よりも後方を探索し、最も近い音符を割り当てる。上記のルールには改善の余地があり、目視によりおよそ正しく動作することを確認したが、全てを確認するに至っていないため、今後修正が必要な可能性がある。

図 7 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、単一音素で出現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、All 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（そもそも歌詞に/a/が多ければ、それぞれのテクニックでも/a/が出現しやすい）。また、テクニック間で分布が異なるのは、音符の高さや長さなどに影響を受けている可能性もある。今後、音素の頻度で正規化した、より正確な分布を求めた上で詳細な分析を行いたい。

また、メリスマは複数音素かつ音高変化を伴うことから一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、ブレイジー、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が多かった。

#### 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毎に楽譜の MIDI ノートナンバーを集計した分布を図 8 に示す。図 8 では、声域の違いを正しく評価するために歌手の性別で分けて可視化した。楽曲の音域の影響も除くことを意図して、歌手毎の音高を正規化して集計した分布を図 9 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、以下のよう求めた。

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

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 8）。その他、しゃくり

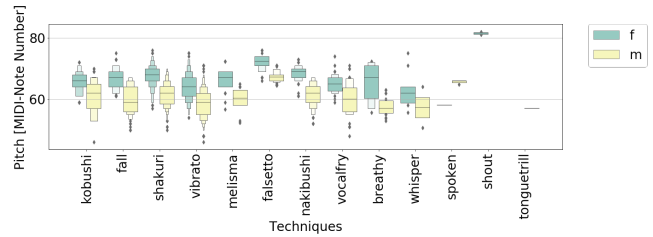


図 8 各歌唱テクニックに対応した音符の音高分布

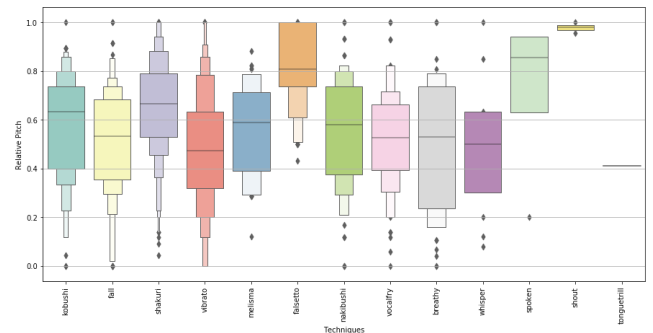


図 9 各歌唱テクニックに対応した音符の音高分布（正規化後）

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印象を受けたが、今後、より詳細な分析が必要である。男性のセリフ語りの音高が高い事例は原曲が「桑田佳祐/勝手にシンドバッド」であり、女性のセリフ語りの音高が低い事例は原曲が「chara/大切をきずくもの」であった。

#### 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 10、図 11 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から、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。最後に、ボーカルフライとヒーカップ (nakibushi) は音の先頭に出現することが多いと考えられるが、フレーズ区切りを自動推定したため（次節で説明）、誤推定の可能性を考えて今回は考慮しなかった。今後の検討課題である。

#### 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 12 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 2 に示した実際の長さとの関係することが分かる。

#### 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出現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出現するか、それ以外の音符で出現するかを集計した。その分

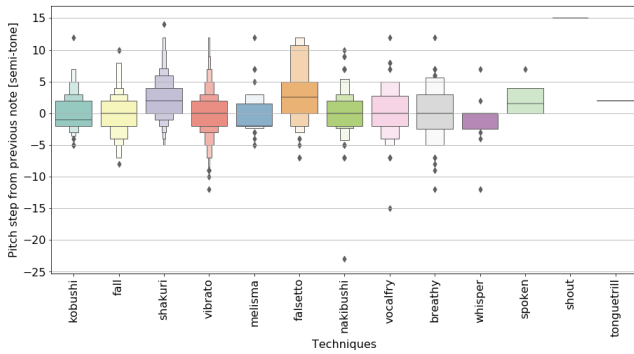


図 10 前の音符からの音高差ごとの分析

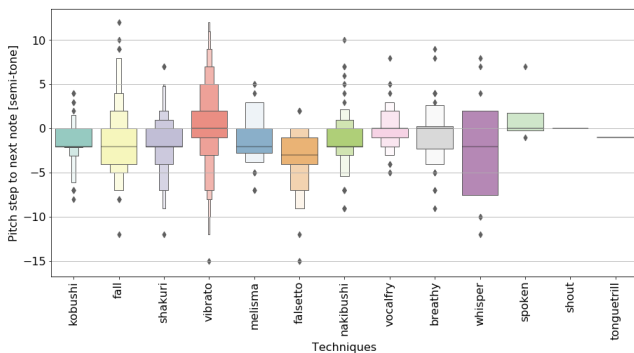


図 11 次の音符への音高差ごとの分析

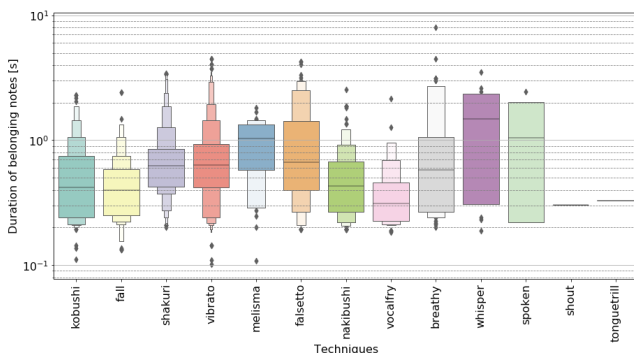


図 12 音符の時間長ごとの分析

布を図 13 に示す。本稿では、16 分音符より長い休符を含む区間をフレーズ境界として、個々のフレーズに自動分割した。歌唱テクニックラベ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り、直感的な印象に合っていた。今後は、フレーズ境界の決定方法も含めて、より分析を深めていきたい。

#### 4.3.6 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、音高及びビブラート長（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。ここで、相関係数はピアソンの積率相関係数にて算出

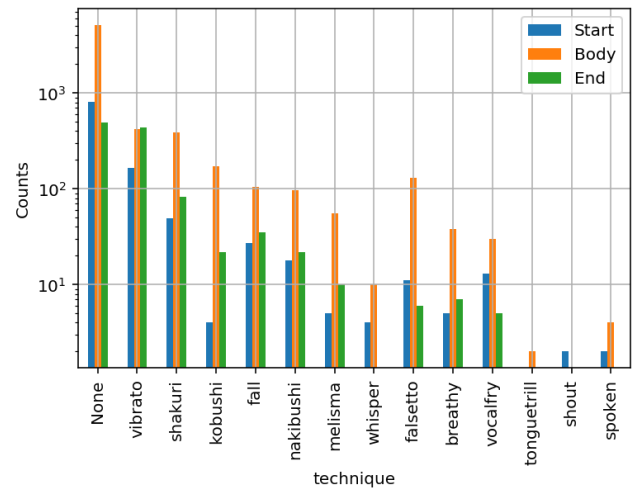


図 13 フレーズの先頭の音符（図中の Start）、末尾の音符（End）、それ以外の音符（Body）の頻度分布。縦軸は対数軸。

表 4 ビブラートパラメータと音高、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）

	音高（女性歌手）	音高（男性歌手）	正規化後の音高	ラベル長
extent	-0.424**	-0.336**	-0.31**	-0.127**
rate	0.225**	0.169**	0.025	-0.268**

した結果を表 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。

## 5. おわりに

本研究ではポピュラー音楽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。テクニックを網羅的に分析し、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、フレーズ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。

## 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. ISMIR 2018*, pp. 514–520 (2018).
- [2] 齋藤 毅, 榎原健一: 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).

- [3] Sundberg, J.: *The Science of the Singing Voice*, Northern Illinois University Press (1987).
- [4] Seachore, C. E.: *A Musical Ornament, the Vibrato, Psychology of Music*, McGraw-Hill (1938).
- [5] Prame, E.: Measurements of the Vibrato Rate of Ten Singers, *J. Acoust. Soc. Am.*, Vol. 96, No. 4, pp. 1979–1984 (1994).
- [6] Prame, E.: Vibrato Extent and Intonation in Professional Western Lyric Singing, *J. Acoust. Soc. Am.*, Vol. 102, pp. 616–621 (1997).
- [7] 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).
- [8] 中野倫靖, 後藤真孝, 平賀 讓: 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).
- [9] 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).
- [10] Ikemiya, Y., Itoyama, K. and Okuno, H. G.: Transferring Vocal Expression of F0 Contour Using Singing Voice Synthesizer, *Proc. IEA/AIE 2014*, pp. 250–259.
- [11] 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. Interspeech 2015* (2015).
- [12] Yang, L.: Computational Modelling and Analysis of Vibrato and Portamento in Expressive Music Performance, PhD Thesis, Queen Mary University of London (2017).
- [13] 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).
- [14] Gupta, C., Li, H. and Wang, Y.: Perceptual Evaluation of Singing Quality, *Proc. APSIPA ASC 2017*, pp. 577–586 (2017).
- [15] Gupta, C., Li, H. and Wang, Y.: A Technical Framework for Automatic Perceptual Evaluation of Singing Quality, Vol. 7, pp. 1–11 (2018).
- [16] 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, 情報処理学会論文誌, Vol. 52, No. 5, pp. 1910–1922 (2011).
- [17] Lee, Y., Oya, M., Kaburagi, T., Hidaka, S. and Nakagawa, T.: Differences Among Mixed, Chest, and Falsetto Registers: A Multiparametric Study, *Journal of Voice (In Press, Corrected Proof)* (2021).
- [18] 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).
- [19] 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).
- [20] Stoller, D., Dixon, S. et al.: Analysis and Classification of Phonation Modes in Singing, *Proc. ISMIR 2016*, pp. 80–86 (2016).
- [21] Wilkins, J., Seetharaman, P., Wahl, A. and Pardo, B. A.: VocalSet: A Singing Voice Dataset, *Proc. ISMIR 2018*, pp. 468–474 (2018).
- [22] 山本雄也, Nam, J., 寺澤洋子, 平賀 讓: 歌唱テクニックの識別における hand-crafted 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).
- [23] Kim, K. L., Lee, J., Kum, S., Park, C. L. and Nam, J.: Semantic Tagging of Singing Voices in Popular Music Recordings, *IEEE/ACM TASLP*, Vol. 28, pp. 1656–1668 (2020).
- [24] Ohishi, Y., Kameoka, H., Mochihashi, D. and Kashino, K.: A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components, *Proc. INTERSPEECH 2012*, pp. 474–477 (2012).
- [25] 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*, pp. 393–397 (2009).
- [26] Arai, M., Matuba, T., Hashida, M. and Katayose, H.: Revealing the Secret of “Groove” Singing: Analysis of J-pop Music, *Proc. SMC 2016*, pp. 21–26 (2016).
- [27] Panteli, M., Bittner, R., Bello, J. P. and Dixon, S.: Towards the characterization of singing styles in world music, *Proc. IEEE ICASSP 2017*, pp. 636–640 (2017).
- [28] Abeßer, J. and Müller, M.: Fundamental frequency contour classification: A comparison between hand-crafted and CNN-based features, *Proc. IEEE ICASSP 2019*, pp. 486–490 (2019).
- [29] Caro Repetto, R., Gong, R., Kroher, N. and Serra, X.: Comparison of the Singing Style of Two Jingju Schools, *Proc. ISMIR 2015* (2015).
- [30] Miryala, S. S., Bali, K., Bhagwan, R. and Choudhury, M.: Automatically Identifying Vocal Expressions for Music Transcription., *Proc. ISMIR 2013* (2013).
- [31] Mauch, M., Frieler, K. and Dixon, S.: Intonation in Unaccompanied Singing : Accuracy , Drift and a Model of Reference Pitch Memory, Vol. 136, pp. 1–11 (2014).
- [32] Nichols, E., Morris, D., Basu, S. and Raphael, C.: Relationships Between Lyrics and Melody in Popular Music, *Proc. ISMIR 2009*, pp. 471–476 (2009).
- [33] 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).
- [34] 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, 研究報告音楽情報科学 (MUS), Vol. 2009, No. 5, pp. 1–6 (2009).
- [35] Wada, Y., Nishikimi, R., Nakamura, E., Itoyama, K. and Yoshii, K.: Sequential Generation of Singing F0 Contours from Musical Note Sequences Based on WaveNet, *Proc. APSIPA ASC 2018*, pp. 983–989 (2018).
- [36] 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. APSIPA ASC 2018*, pp. 1003–1009 (2018).
- [37] 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年9月現在, 未公判) (2021).
- [38] Cannam, C., Landone, C. and Sandler, M.: Sonic visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files, *Proc. ACM MM 2010*, pp. 1467–1468 (2010).
- [39] Kim, J. W., Salamon, J., Li, P. and Bello, J. P.: CREPE: A Convolutional Representation for Pitch Estimation, *Proc. IEEE ICASSP 2018*, pp. 161–165 (2018).
- [40] Saitou, T. and Goto, M.: Acoustic and Perceptual Effects of Vocal Training in Amateur Male Singing, *Proc. INTERSPEECH 2009*, pp. 832–835 (2009).