

小特集—話す・歌う・奏でる音の合成技術—

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——*

中野倫靖, 後藤真孝 (産業技術総合研究所)**

43.75.Rs

1. はじめに

歌声には多様な歌唱表現があり, 歌声を自動的に合成する際には, そうした表現をいかに制御するかが重要となる。テキスト音声合成でも感情表現が扱われることがあるが, 歌声合成では音楽的な表現ができることが本質的である点が大きく異なる。本稿では, そうした歌声合成で必須となる基盤技術として, 多様な表現を得るための特徴量の制御について述べる。歌声合成では, 合成方式や歌手ごとの音源 (ある特定の歌手の歌声を合成するためのデータ集合) の構築方法も重要だが, 本稿では表現の制御のみに焦点を当てる。具体的には, 歌声合成における歌唱表現制御に関する Umbert *et al.* による解説 [1] を, より最近の研究や, Human-computer interaction (HCI) の観点からの表現制御を含めて発展させる形で解説する。

歌詞の意味を伝えつつ音楽的な表現ができる歌声は, 音楽の重要な要素であり, 歌声や歌詞を中心に音楽を聴く人は多い [2]。音楽制作においても中心的に扱われる。更に, 歌声は音の三要素である音高・音量・音色のすべてがリズムを伴って複雑に時変化する時系列信号であり, 特に音色は歌詞の音韻が次々と変化するため, 他の楽器音の生成・制御よりも技術的に難易度が高い。このような歌声の多様な歌唱表現を効果的で自然に制御できる技術や, 直感的で容易に制御するためのユーザインタフェースの実現は, 学術的にも意義がある。

音楽制作において, 歌声合成技術は既に広く利用されている。特に日本を中心に 2007 年以降, 市販の歌声合成ソフトウェアである VOCALOID2

「初音ミク」等が注目を集め, 歌声合成を用いた楽曲が数十万曲制作された。それらは音楽ヒットチャート上位に登場する等, 楽しみリスナも増加した [3, 4]。歌唱曲を制作する敷居を歌声合成が下げたため, 一般の人々も楽曲制作に参加して 2 次創作の活発な文化を育み, 大規模な協調的創造活動を引き起こした [5, 6]。そうした文化・活動の中で生まれた楽曲は, 歌声合成に基づくキャラクターがコンピュータグラフィクス (CG) の映像でステージに登場する先進的なライブコンサートで演奏されるようになった [7]。国内外で毎年開催され続けている。このように音楽創作・鑑賞における可能性を広げてきた歌声合成は, 多様な音楽ジャンルで活用されており, より一層多様な歌唱表現の制御ができる技術の需要が高まっている。

2. 歌声の可能性を広げる歌声合成技術

歌声合成は, 人間の多様な歌唱を再現して合成するだけでなく, 人間には不可能な歌唱すら合成できる可能性を持っている。音楽制作においても人間の歌唱と歌声合成は必要に応じて使い分けられたり, 相補的に活用されたりすることが重要である。そこで, 多様な歌唱表現を制御して歌声合成の可能性を広げることを念頭に, 本章では, 人間と歌声合成のそれぞれの能力・特長と限界を整理し, 歌唱表現制御のためにどのような技術が求められているかを議論する。

2.1 人間の歌声生成 (歌唱) 能力とその限界

人の歌声は, 人間らしくて自然性が高く, 歌唱力があれば音楽的に多様な表現を持つ歌声を生成できる特長を持つ。また, 歌詞に応じた多様な歌い回しを臨機応変に駆使できる表現力も持つ。

しかし, 過去に歌った歌を完全に再現してもう一度歌ったり, 自分の歌唱力を超えた歌を歌ったりすることは難しい。例えば, 自身の限界よりも広い声域の歌を歌ったり, 歌詞が速すぎる歌を歌つ

* Fundamental technology for singing synthesis: Control of acoustic features.

** Tomoyasu Nakano and Masataka Goto (National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, 305-8568) e-mail: t.nakano@aist.go.jp; m.goto@aist.go.jp

たりすることはできない。

2.2 歌声合成の能力とその限界

歌声合成は、多様な声質の歌声を再現性高く何度でも合成できる特長を持つ。楽曲制作者の立場からは、歌声合成ソフトウェアを用いればいつでもどこでも歌声が得られ、人間の歌唱者と違ってソフトウェアは疲労しないために、表現を少しずつ変えながら何度でも調整できる。また、歌声を音高や音量のような個別の要素に分解して制御でき、複数の声色（音源）を切り替えたり混ぜ合わせたりできる（市販の歌声合成ソフトウェア VOCALOID4 のクロスシンセシス機能 [8] 等）。

しかし、人間の歌声と区別がつかないような自然な歌声を、人間と同じくらい多様な表現力で自動的に生成できる技術は、まだ実現されていない。近年、歌声合成音の自然性は飛躍的に向上したが、歌唱表現の多様さと歌詞の意味に応じた表現力は人間には及ばない。更に、ライブステージや即興演奏で、リアルタイムかつ完璧に歌唱表現を制御するのは困難である。特に、楽器音の生成とは異なり、音高・音量に加えて歌詞もその場で指定する必要があると一層困難となる。

2.3 歌声合成に求められる要件

前節でまとめた歌声合成の能力や限界に基づき、歌唱表現制御に求められる要件を議論する。歌声合成には、以下に挙げるような目的があるため、それを達成するための技術を要する。

- 1) 自然な歌を合成する
 - 2) 多様な声質・声色で合成する
 - 3) 多様なスタイル・テクニックで合成する
 - 4) リアルタイムに歌声合成・表現制御をする
- 1) を実現する上では、歌詞と楽譜（メロディ）を入力として、自然な音韻・音高・音長（音価）等で歌声が合成できるとよい。2) においては、様々な歌手の声質や、同じ歌手でも多様な声色で合成する必要がある。例えばデスボイス等の特殊な技能を要する表現も合成対象となる。3) では、ビブラート等の歌唱特有のテクニックの制御が求められる。そして4) は、伴奏や背景音楽にその場で同期させた制御を必要とする。

以上の目的のために、音の三要素である音高、音量、音色に加え、歌詞の音韻やリズム（[1] の Table 3 参照）を歌唱表現の要素として制御しなければならない。音高は基本周波数（ f_0 ）、音量は波

形振幅、音色はスペクトル包絡形状（共振周波数）等により制御する。歌詞は文字列や音素列のテキストとして与えることが多く、リズムは歌詞に対応する音符の時刻や音価、無音の挿入等によって制御する。その上で、リアルタイム制御のためには、いかに低いレイテンシ（遅延時間）で合成できるかも重要となる。

3. 制御方法

2.3 節において議論した要件に対し、その制御方法は多様である。本章では、既存の研究事例を以下の6種類の観点で分類して紹介する。

- ピアノロールなどの楽譜表現を用いる（3.1 節）
- お手本の歌声を真似る（3.2 節）
- 歌唱テクニックを指定する（3.3 節）
- 既存楽器を用いる（3.4 節）
- 身体動作を用いる（3.5 節）
- 合成された歌声を変形する（3.6 節）

これらの観点は独立ではなく、複数が混在する場合もある。

3.1 歌を楽譜で制御する

市販の歌声合成ソフトウェアでは、歌い方を、ピアノロールと呼ばれる楽譜表現（音符系列）で制御する方式が一般的である。歌詞テキストとピアノロールを入力として歌声を合成することから、話声合成の text-to-speech synthesis と対応付けて text-to-singing synthesis もしくは lyrics-to-singing synthesis と呼ばれることがある [9]。ピアノロール自体は、ユーザが MIDI キーボード（ピアノ鍵盤）やマウス等を用いて入力し、そこに歌詞のテキストを文字入力用キーボード等から順次打ち込んで対応付けることが多い。音高・音量の細かな変動や声色変化を与えるパラメータは、事前に用意されたテンプレートに基づいて制御したり、マウス等で個別に入力して制御したりできる。この方式は、VOCALOID [10] や HMM 歌声合成システム [11] 等で広く採用され、近年の深層学習に基づく歌声合成システム [12, 13] でも前提とされている。

様々な歌手の歌声や異なる声色（音色）を合成するためには、別々の音源を用意して使い分けることが多い。あるいは、共振周波数や非周期成分の度合い等を変更することで、音色制御を実現することもある。特殊な事例としては、話し声を歌

声に変換する speech-to-singing synthesis と呼ばれる方式があり、人間の話し声の声質を利用しつつ、話し声の声の高さを変更して音素を引き延ばすことで歌声にするシステム [9,14] も提案されている。

3.2 歌を歌で制御する

人の歌声をお手本として与えてその歌い方を真似て歌声合成をする方式を singing-to-singing synthesis と呼ぶ [15]。このような「真似る」アプローチは、複雑で時間のかかる楽譜入力や手作業によるパラメータ調整をしなくても、「このように歌わせたい」というお手本の歌声と、その歌詞テキストを与えるだけで利用できる利点がある。

Janer *et al.* [16] は、お手本の歌声から推定した音高・音量・ビブラート情報（深さ・速さ）を正規化し、歌声合成パラメータへ変換する手法を提案した。中野ら [15] は、お手本の歌声から推定した音高・音量を用いて既存の歌声合成システムで合成しても、必ずしもお手本どおりにならない問題に着目し、その合成した歌声の音高・音量がお手本に近づくように反復しながら自動調整できるシステム VocaListener を実現した。これは、歌声合成システムの違いや音源の違いに対処できるだけでなく、歌声分析結果の誤りをユーザがインタラクティブに訂正できる機能も有していた。音高のずれやビブラート等の歌唱表現を修正できる支援機能により、お手本の歌唱力を超えた表現を得ることもできた。更に、同一歌手による複数の声色の音源を活用することで、お手本の声色変化を真似る拡張もなされた [17]。

上記の方法では歌詞のテキストを事前に用意する必要があるが、実用性があり、VocaListener は市販されていた。将来的には、歌詞も自動認識しながらお手本を真似る技術が登場する可能性がある。歌詞を伴わないハミングのようなお手本歌唱も真似る場合には、楽器の演奏を真似る InstListener [18] のようなアプローチも考えられる。

3.3 歌を歌唱テクニックで制御する

どのような歌唱テクニックをどこで用いるかを、音符系列やその歌詞に応じた的確に制御することは、個性的な歌唱スタイルの表現や歌唱力の高い歌声の合成につながる。例えば、歌声の自然性に寄与する歌声特有の音響特徴量として、オーバシュート（音高変化直後の目的音高を越える瞬時的な変

動成分）やビブラート等の f_0 動的変動成分 [19] があるが、それらの含まれ方は歌唱テクニックや歌唱力によって異なる。例えば、オーバシュートは歌唱力の差によらず生じるが、その変動の大きさは歌唱力に依存し、プロによる歌唱の方がアマチュアによる歌唱よりも変動が小さい [20] ことから、歌唱力の高さを感じさせるためには、その変動量を小さくする必要がある。また、ビブラートの深さ（変動幅）は歌手によって異なることから [21]、その制御が歌唱スタイルに影響を与える。

歌唱テクニックを指定した表現制御では、VOCALOID におけるビブラート等のプリセット選択や手書きによる制御の他、グロウル [8,22] や、シャウト [23]、グイン [24]、長唄特有の「あたり」と「ふり」 [25] のような表現を個別に付与する手法が提案されてきた。また、歌唱テクニックに着目したデータセットも整備され始めており、その表現制御のための研究開発に有用である。例えばアカペラ歌唱のデータセット VocalSet [26] には、ビブラートやトリル等の 16 種類の歌い方（それに話し声を加えた合計 17 種類）が収録されている。

歌唱テクニックや歌唱力に関する制御は、歌の印象を大きく変える効果があって重要であり、上述のような個別の取り組みでも実用性はある。将来的には、多様な歌唱テクニックを総合的に扱えるような研究開発が進展することで、より一層豊かな表現制御が実現されることが期待できる。

3.4 歌を楽器で制御する

上述の歌唱表現制御ではいずれもリアルタイム性が考慮されていないが、リアルタイムに制御する目的では、MIDI キーボードのような既存楽器が用いられることが多い。しかし、歌声の歌詞（音韻）とメロディ（韻律）とを同時に入力するのは容易でなく、様々な工夫が必要となる。

まず、左手で歌詞を、右手でメロディを同時に入力するリアルタイム歌唱合成インタフェースが提案された。山本ら [27] は、左手では文字入力用の専用インタフェースで歌詞の各文字の子音と母音を同時に指定し、右手では MIDI キーボードの鍵盤を演奏することで音高を入力するインタフェースを提案した。一方、フォルマン兄弟は、MIDI キーボードのみを用いてその音域を左右に分け、左手側の 1 個～2 個の鍵盤の組み合わせで歌詞の各文字（ひらがな）の子音と母音を同時に指定し、右手

側の鍵盤で音高と音量を入力するインタフェースを提案した [28]。インタフェースとしては、MIDI アコーディオンも用いられていた。

次に、歌詞を事前に与えておくことで、入力を容易にするインタフェースも提案された。Yamamoto *et al.* [29] は、歌詞のひらがな文字列を事前登録した上で、音域を左右に分けた MIDI キーボードの右手側では音高を指定し、左手側では 1 個の鍵盤を押して母音だけを指定する入力インタフェースを提案した。母音系列を歌詞とマッチングすることで多少の誤りが含まれていても歌声が合成できる。他にも、歌詞を事前登録し、右手で音高のみを入力する VOCALOID Keyboard が市販されている。

一方、歌声合成とは異なるが、歌声の音高を楽器で制御する研究がある。Morikawa *et al.* [30] は、喉頭摘出者が電気喉頭を用いて発声した音高変化が乏しい音声を、鍵盤楽器の演奏音の音高に基づいて歌声に変換できる歌唱支援システムを提案した。また、Ojima *et al.* [31] は、既存の楽曲の混合音中に含まれるボーカルの歌声の音高を、MIDI キーボードによってリアルタイムに変更して演奏できるインタフェースを提案した。

3.5 歌を身体動作で制御する

歌唱表現のリアルタイム制御では、上述の楽器を用いる方法だけでなく、様々な身体動作を直接利用できることができれば便利である。楽器に不慣れた人が直感的に歌唱表現を制御したり、歌詞入力とメロディ入力のモダリティを分ける（手とそれ以外を併用する）ことで同時入力の困難さを減らしたりできるからである。ただし、習熟には長時間を要するインタフェースもあり、まだ普及はしていない。

Fels *et al.* [32] は、ジェスチャーによって歌声と話し声を合成できる手袋型のデバイスを提案した。右手の前後左右の位置で母音、右手の指の動きで子音、左手の上下位置で音高を制御できる。Ito *et al.* [33] は、音高、音量、ビブラートの深さを腕の動きで、母音の種類を手首のひねりで操作するシステムを提案した。竹本らは、鼻歌による音高入力と、文字入力キーボードによるローマ字歌詞入力を組み合わせ、リアルタイムに歌声を合成できるインタフェース「HANAUTAU」 [34] と、歌詞に点字入力を用いたトランペット型歌声

合成インタフェース [35] を提案した。

また、歌詞を伴わないが、D'Alessandro *et al.* [36] は、ペン型のデバイスによって、音高・音量等をリアルタイムに制御する歌声合成インタフェースを提案した。安ら [37] は、微小な無声音と咽頭位置をセンシングすることで、発声なしに鼻歌の音高・音量を再現するシステムを提案した。

他にも、森勢ら [38] は事前に用意された歌詞と楽譜情報を用いて、タブレットの画面を指でなぞることでビブラートの深さと速さを変更できるインタラクションを提案した。Hewitt *et al.* [39] は、人間の歌手の歌声のフィルタリング等の加工をジェスチャーをセンシングして行うマイクスタンド eMic を提案した。また、Park *et al.* [40] は、マイクに内蔵されたセンサにより、歌手のジェスチャーを使用してビブラート生成や音高制御、エフェクト付与を可能とするインタフェースを提案した。

3.6 歌を加工・変形して制御する

歌唱表現の制御という観点からは、人間の歌声を加工・変形するために開発された技術を、そのまま合成された歌声にも適用することができる。例えば、Auto-Tune 等の市販の音高補正用ソフトウェアは、人間の歌声の音高やタイミングのずれを補正するために、商業音楽制作において広く用いられている。これは、人間が発声できないような機械的な音高変化を付与するために用いられることがあり、合成した歌声を敢えて機械的に加工するために適用されることがあるのは興味深い。

ある人の歌い方を別の人の歌い方に変形する技術や、それらの中間の歌声を生成するモーフィング技術も活発に研究開発されてきた。そうした技術を適用すれば、合成された異なる歌唱表現の歌声同士をモーフィングしたり、合成された歌声と人間の歌声とをモーフィングしたりすることが可能である。

4. 今後の展望

本稿では、歌声合成における歌唱表現を、楽譜・歌声・歌唱テクニック・楽器・身体等の様々な手段で制御する研究事例を紹介した。歌唱表現制御の研究開発は発展途上で様々な可能性があり、チャレンジングで未解決な問題も多く、魅力的な研究テーマの宝庫である。

歌声合成や歌唱表現制御の研究開発は今後、人間の歌声と区別できないくらい限りなく人間の歌声に近づいていく方向性と、逆に、人間には不可能な歌唱表現を生み出していく方向性のそれぞれを極める形で発展していくと我々は予想している。その両者について、4.1 節と 4.2 節で議論する。

歌声合成がより一層普及するためには、利用場面を拡大するための研究開発も不可欠である。人間の優れた歌手が持つ多種多様な表現を誰でも手軽に思いのままに合成できるようにする技術や、ライブコンサートで観客の反応に応じて歌唱表現を変えられるぐらいのリアルタイム性と柔軟性を生み出す技術は、まだほとんど研究開発がなされていない。そうした展望を、4.3 節と 4.4 節で議論する。

4.1 限りなく人間に近い歌声の合成

合成された歌声は、今後ますます自然になり、多様な表現が可能になって、最終的には人間の歌声と区別ができなくなっていく。人間の歌声の表現力は幅広く、ある一つの表現では自然性を達成できても、あらゆる表現で自然な歌声を合成することは難しい。あらゆる歌手の個性、声質・声色、歌唱力を再現することも難しく、いかにそれらを制御可能にしていくかが、歌唱表現制御におけるグランドチャレンジとなる。

4.2 人間には不可能な歌声の合成

合成された歌声は、人間のような喉という物理的な制約がないため、人間の歌声とは違う独自の歌声と表現を生み出し、多様な音楽の創造に寄与していく。これは、楽音合成（シンセサイザ）が自然な楽器音だけでなく、それ以前には存在しなかった楽器音と表現を生み出してきたのと同様である。人間には困難な高速な歌詞や声域の広い歌声は、歌声合成を用いた様々な楽曲で既に利用されているが、聞いたことがない未知の歌声をどうデザインし、どう合成すればよいのかは明らかでない。どこまでが歌声なのか、楽器音との境界は何なのか、といった様々な議論が起きることが予想される。

4.3 多種多様な歌声を自由自在に合成

あたかもプロの歌手が自分のために歌ってくれるかのように、将来は、歌詞に即した適切で多様な歌唱表現が、時間をかけずに手軽に思いどおりに合成可能になっていく。歌声合成を用いた楽曲

制作で現在最も普及しているのは 3.1 節で述べた方式だが、所望の表現を得るためには多くの時間と労力がかかる。音楽経験の多寡を問わず誰でも使いこなせて表現の自由度が高いインタラクティブな、歌詞の意味を反映した歌唱表現制御は難しく、それらの実現はグランドチャレンジとなる。

4.4 歌声のリアルタイムで柔軟な合成

ライブコンサートで人間の歌手が観客の反応に応じて歌唱表現を変えるように、将来は、歌声合成が同じフレーズでもリアルタイムに歌い方を変えながら表現できるようになっていく。ライブコンサートやイベント等での歌声合成の利用は拡大したが、現在は事前に合成した歌声を再生するのが一般的である。リアルタイムで動的に変わる歌唱表現制御の実現は難しく、しかもそれを人間がインタラクティブに制御できるだけでなく、上記の観客の反応のようなその場にならないと判明しない状況に応じて制御できる技術は、実現の目処が立っていない。こうしたリアルタイム性と柔軟性を持つ歌声合成は、今後、様々な発展の可能性がある。

文 献

- [1] M. Umbert, J. Bonada, M. Goto, T. Nakano and J. Sundberg, "Expression control in singing voice synthesis: Features, approaches, evaluation, and challenges," *IEEE Signal Process. Mag.*, 32, 55–73 (2015).
- [2] A. Demetriou, A. Jansson, A. Kumar and R. M. Bittner, "Vocals in music matter: The relevance of vocals in the minds of listeners," *Proc. ISMIR 2018*, pp. 514–520 (2018).
- [3] H. Kenmochi, "VOCALOID and Hatsune Miku phenomenon in Japan," *Proc. InterSinging 2010*, pp. 1–4 (2010).
- [4] 後藤真孝, "初音ミク, ニコニコ動画, ピアプロが切り拓いた CGM 現象," 情報処理 (情報処理学会誌), 53, 466–471 (2012).
- [5] 濱野智史, "インターネット関連産業," デジタルコンテンツ白書 2009 (デジタルコンテンツ協会, 東京, 2009), pp. 118–124.
- [6] M. Hamasaki, M. Goto and T. Nakano, "Songrium: A music browsing assistance service with interactive visualization and exploration of a web of music," *Proc. WWW 2014*, pp. 523–528 (2014).
- [7] 伊藤博之, "初音ミク as an interface," 情報処理 (情報処理学会誌), 53, 477–483 (2012).
- [8] 橋 誠, "歌声合成ソフトウェア VOCALOID4 における表現力向上への取り組み," 情報研報, Vol. 2015-MUS-107, No. 59, pp. 1–6 (2015).
- [9] T. Saitou, M. Goto, M. Unoki and M. Akagi, "Speech-to-singing synthesis: Converting speaking voices to singing voices by controlling acoustic features unique to singing voices," *Proc. WASPAA 2007*, pp. 215–218 (2007).
- [10] H. Kenmochi and H. Ohshita, "VOCALOID—

- Commercial singing synthesizer based on sample concatenation,” *Proc. Interspeech 2007*, pp. 4009–4010 (2007).
- [11] K. Oura, A. Mase, T. Yamada, S. Muto, Y. Nankaku and K. Tokuda, “Recent development of the HMM-based singing voice synthesis system—sinsy,” *Proc. SSW7*, pp. 211–216 (2009).
- [12] M. Nishimura, K. Hashimoto, K. Oura, Y. Nankaku and K. Tokuda, “Singing voice synthesis based on deep neural networks,” *Proc. Interspeech 2016*, pp. 2478–2482 (2016).
- [13] M. Blaauw and J. Bonada, “A neural parametric singing synthesizer modeling timbre and expression from natural songs,” *Appl. Sci.*, 7, 1–23 (2017).
- [14] K. Vijayan, H. Li and T. Toda, “Speech-to-singing voice conversion: The challenges and strategies for improving vocal conversion processes,” *IEEE Signal Process. Mag.*, 36, 95–102 (2019).
- [15] 中野倫靖, 後藤真孝, “VocaListener: ユーザ歌唱の音高および音量を真似る歌声合成システム,” 情処学論, 52, 3853–3867 (2011).
- [16] J. Janer, J. Bonada and M. Blaauw, “Performance-driven control for sample-based singing voice synthesis,” *Proc. DAFx-06*, pp. 41–44 (2006).
- [17] 中野倫靖, 後藤真孝, “VocaListener2: ユーザ歌唱の音高・音量に加えて声色変化も真似る歌声合成システム,” 情処学論, 54, 1771–1783 (2013).
- [18] Z. Shi, T. Nakano and M. Goto, “InstListener: An expressive parameter estimation system imitating human performances of monophonic musical instruments,” *Proc. ICASSP 2018*, pp. 581–585 (2018).
- [19] 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,” 音響学会誌, 64, 616–623 (2008).
- [20] 齋藤 毅, 鶴木祐史, 赤木正人, 榊原健一, “歌声の基本周波数変化に含まれるオーバーシュートの知覚への影響に関する検討,” 音響学会聴覚研資, H-2006-109, pp. 611–616 (2006).
- [21] 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, 137, 1607–1614 (2017).
- [22] J. Bonada and M. Blaauw, “Generation of growl-type voice qualities by spectral morphing,” *Proc. ICASSP 2013*, pp. 6910–6914 (2013).
- [23] Y. Nishigaki, K. Sakakibara, M. Morise, R. Nisimura, T. Irino and H. Kawahara, “Controlling “shout” expression in a Japanese POP singing performance: Analysis and suppression study,” *Proc. Interspeech 2013*, pp. 2905–2909 (2013).
- [24] 森勢将雅, 村主大輔, 馬場 隆, 片寄晴弘, “奄美大島民謡風の歌唱デザインを支援するシステム: グインレゾネータ,” 情処学論, 54, 1244–1253 (2013).
- [25] N. Minematsu, B. Matsuoka and K. Hirose, “Prosodic analysis and modeling of NAGAUTA singing to generate prosodic contours from standard scores,” *IEICE Trans. Inf. Syst.*, E87-D, 1093–1101 (2004).
- [26] J. Wilkins, P. Seetharaman, A. Wahl and B. Pardo, “VocalSet: A singing voice dataset,” *Proc. ISMIR 2018*, pp. 468–474 (2018).
- [27] 山本和彦, 加々見翔太, 濱野桂三, 柏瀬一輝, “リアルタイム日本語歌唱鍵盤楽器のための文字入力インタフェースの開発,” 情処学論, 54, 1373–1382 (2013).
- [28] 佐近田展康, “「兄弟式リアルタイム音声合成演奏システム」の概要と背景,” 名古屋学芸大学メディア造形学部研究紀要, 6, 21–33 (2013).
- [29] K. Yamamoto and T. Igarashi, “LiVo: Sing a song with a vowel keyboard,” *J. Inf. Process.*, 24, 460–468 (2016).
- [30] K. Morikawa and T. Toda, “Electrolaryngeal speech modification towards singing aid system for laryngectomees,” *Proc. APSIPA ASC 2017*, pp. 610–613 (2017).
- [31] Y. Ojima, T. Nakano, S. Fukayama, J. Kato, M. Goto, K. Itoyama and K. Yoshii, “A singing instrument for real-time vocal-part arrangement of music audio signals,” *Proc. SMC 2017* (2017).
- [32] S. S. Fels and G. E. Hinton, “Glove-TalkII: A neural-network interface which maps gestures to parallel formant speech synthesizer controls,” *IEEE Trans. Neural Networks*, pp. 205–212 (1998).
- [33] M. Ito, T. Ashina and Y. Saegusa, “A singing voice synthesizer controlled by arm motions using compressed phoneme determination algorithm,” *Proc. IIH-MSP 2014*, pp. 550–553 (2014).
- [34] 竹本拓真, 馬場 隆, 片寄晴弘, “リアルタイムに初音ミクを歌わせるタイプソングシステム「HANAUTAU」とそのアジャイル型開発事例報告,” インタラクシオン 2014, pp. 682–685 (2014).
- [35] 竹本拓真, 馬場 隆, 服部篤志, 片寄晴弘, “VOCALUMPET: トランペット型リアルタイム歌声シンセサイザ,” エンタテインメントコンピューティング 2014, pp. 75–76 (2014).
- [36] N. D’Alessandro and T. Dutoit, “HandSketch bi-manual controller: Investigation on expressive control issues of an augmented tablet,” *Proc. NIME 2007*, pp. 78–81 (2007).
- [37] 安謙太郎, 竹内祐太, 稲見昌彦, “サイレントハミング: 呼吸音と咽頭運動を利用した鼻歌再現システム,” エンタテインメントコンピューティング 2011, pp. 1–2 (2011).
- [38] 森勢将雅, 渡邊優介, “実時間音声合成技術を活用した歌唱デザインインタフェースの試作,” 情処研報, Vol. 2019-MUS-122, No. 30, pp. 1–6 (2019).
- [39] D. Hewitt, “eMic: Developing works for vocal performance using a modified, sensor based microphone stand,” *Proc. CHI EA ’13*, pp. 2943–2946 (2013).
- [40] Y. Park, H. Heo and K. Lee, “VOICON: An interactive gestural microphone for vocal performance,” *Proc. NIME 2012*, pp. 43–46 (2012).

中野 倫靖

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。情報処理学会, 日本音響学会各会員。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

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情報技術研究部門 首席研究員 兼メディアコンテンツ生態系プロジェクトユニット代表。JST ACT-I「情報と未来」研究総括, 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, 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, 47件受賞。