

解説

歌声情報処理の最近の研究*

後藤真孝, 齋藤 毅, 中野倫靖, 藤原弘将 (産業技術総合研究所)**

43.75.-z

1. 注目を集める歌声情報処理

近年、音楽情報処理分野の発展と共に [1-3], 歌声に関する研究活動が世界的に活発に取り組み、学術的な観点からだけでなく、産業応用的な観点からも注目を集めている。そうした歌声に関する研究は、歌声固有の特徴に関する基礎研究から、歌声合成、歌詞認識、歌手同定、歌声検索、歌声評価等の応用研究まで多岐に渡る。そこで本解説では、そうした歌声に関する幅広い研究を「歌声情報処理」と名付け、具体的な研究事例をできるだけ多く挙げて紹介する。

歌声は、音声と音楽の両方の側面を持つが、いずれの分野の観点からも未解決の研究課題は多い。例えば、歌声は通常の音声よりもはるかに変動が大きく、また、歌声と相互に関連し合う伴奏音も大きな音量で含まれているため、その自動認識は、技術的に最も難しいクラスの音声認識問題であると言える。実際、歌声の歌詞の自動認識は、まだほとんど実現できていない。音楽の認識・理解の観点からも、従来主に研究されてきた楽器音に比べ、歌声の変動の大きさは並外れており、技術的に難しくかつ興味深い課題が多い。歌声の合成に関しても、話声のように言語として内容が伝わる必要があることに加え、声の高さや強さの動的で複雑な変化や歌声としての表現力が求められ、まだまだ研究途上で課題も多い。このように歌声情報処理の研究は、学術的にも正にフロンティアである。

その上、音楽は産業・文化の面で主要なコンテンツの一つであり、歌声は音楽の最も重要な要素であることから、その研究成果は社会的にも大きなインパクトを持っている。既に商業音楽（特に

ポピュラ音楽) の制作では、歌声の音高を信号処理で補正する歌声情報処理技術が日常的に用いられており、歌手の歌唱力が不十分な箇所の音程補正や意図的な表現効果を狙った補正等で、必要不可欠なものとなっている。歌声を用いた楽曲検索も実用化され、メロディを歌ったりハミングしたりすると曲名が分かるサービスは、携帯電話で簡単に利用できる。最近では、歌声の合成も注目を集め、歌声合成技術を用いて制作された楽曲が、ニコニコ動画や YouTube のような動画共有サービスに大量に投稿されている。まだ実験的ではあるが、そうした歌声合成技術を用いた楽曲を収録した音楽 CD すら市販されている。ほかにも、カラオケの歌声採点機能等が普及している。特にポピュラ音楽では歌声を中心に音楽を聴く人達が多く、今後も、歌声に関する様々なアプリケーションの登場が期待されている。

本解説では、通常の歌唱以外にもハミングや口(くち)ドラム(ドラムの擬音語表現)等、人間の口から発声される音楽に関連した声をすべて「歌声」と捉え、この「歌声情報処理」という研究分野を様々な角度から解説していく。

2. 歌声情報処理の発展の背景

近年の歌声情報処理の急速な発展は、長年にわたって取り組まれてきた歌声に関する基礎研究の存在に支えられている。その中から、歌声の音響分析によって明らかとなってきた歌声特有の音響特徴を紹介し、歌声情報処理の研究にとって重要な歌声データベース(DB)について述べる。

2.1 歌声特有の音響特徴

歌声は、話声と比較して発声の高さや強さの変動幅が広く、より複雑で動的な特性を持つことが知られている。特に楽曲の旋律に沿って変化する基本周波数(F_0 , 声の高さ)の軌跡には、歌声固有の動的な変動成分が現れる。中でも、以下に示

* Recent studies on singing information processing.

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

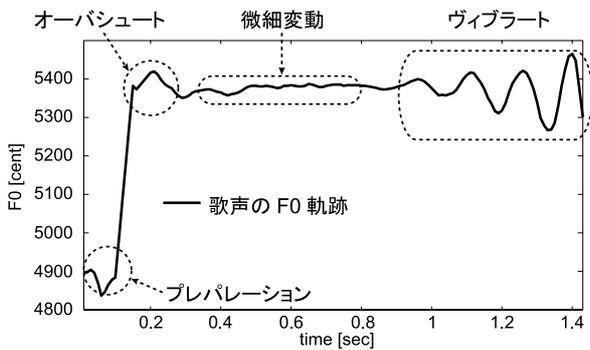


図-1 歌声における 4 種の基本周波数 (F_0) 動的変動成分。

す 4 種の成分は、歌唱法や歌唱者に依存せず様々な歌声において共通して存在し、かつ歌声を知覚する上で重要な役割を担っていることが明らかとなってきた (図-1) [4]。

オーバーシュート：滑らかな音高変化、及びその直後に目標音高を超える瞬時的な変動成分

ヴィブラート：同一音高区間で観測される 4~8 Hz の準周期的な変動成分

プレパレーション：音高変化直前に変化とは逆方向に触れる瞬時的な変動成分

微細変動：発声区間全体に観測される不規則で細かい変動成分

また、スペクトルにおける歌声特有の特徴も存在し、中でも Sundberg によって報告された singer's formant (歌手のホルマント) が代表的である [5]。これは、主に男性オペラ歌唱の母音スペクトルの 3kHz 付近に存在する顕著なピーク成分で、歌声に聴感的な響きを与える要因と考えられており、邦楽 (日本の伝統芸能) 歌唱においても観測されることが報告されている [6, 7]。図-2 は、後述する歌声 DB 「日本語を歌・唄・謡う」 [8] に収録されているテノール歌手による歌声と話声、及び民謡歌手による歌声の母音/a/発声時の長時間平均スペクトルである。歌声の方が話声に比べて、テノール歌唱では 3kHz、民謡歌唱では 4kHz 付近のスペクトル成分がそれぞれ強いことが分かる。そのほかにも歌声特有の音響特徴は多数報告されているが、詳細に関しては参考文献 [5] を参照されたい。

2.2 歌声データベース (DB)

歌声を対象とした研究では、これまで各研究毎に歌唱音声を取録して実施することが多かった。しかし近年、研究上共通利用が可能な、歌声に関する音楽 DB が増えつつある。

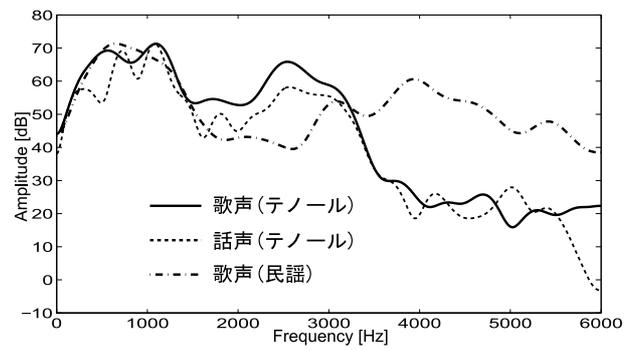


図-2 テノール及び民謡歌唱における singer's formant の一例。

まず、ハミング検索 (6.1 節で後述) の研究に利用する目的で収録された研究用 DB が幾つかある [9-11]。そこでは、ハミング (鼻歌、歌詞なし歌唱) や歌唱、口笛などの音響信号に加え、その歌唱の音高・音長を記録した標準 MIDI ファイルが提供されている。

更に、国内では声楽等のクラシック音楽に加えて、邦楽やポップ音楽などを含めてジャンルのバラエティに富んだ大規模な DB として以下のようなものがある。

RWC 研究用音楽 DB (RWC-MDB) [12] 研究目的で共通利用できる世界初の著作権処理済み音楽 DB である。ポップ音楽、ロック、ジャズ、ラテン、声楽、邦楽、童謡や英語の有名曲、単音の歌唱など、多様なスタイルの歌唱とその歌詞が用意されている。

日本語歌唱音声 DB 「日本語を歌・唄・謡う」 [8] 邦楽 (日本の伝統芸能) のほぼすべてのジャンルと、洋楽の各ジャンルの歌唱が取録された DB である。同一の歌詞をそれぞれの典型的歌唱法で発声した歌唱音声と朗読音声が取録されている。

AIST ハミング DB (AIST-HDB) [13] RWC-MDB の曲を初めて聴く歌唱者が 1 回及び 5 回聴いた後に、思い出しながら歌った音声を収録した DB である。同一メロディの歌唱とハミングに加え、朗読音声も収録されている。まだ未公開であるが、今後の公開が検討されている。

3. 歌う計算機

計算機に歌を歌わせる、すなわち歌声合成は、歌声情報処理の分野では最も古くから取り組まれている研究課題の一つである。そして、この歌声合成の研究は、ここ数年で多様な展開を見せている。

歌声合成の歴史を少し遡ってみると、特に 1980 年代後半から 1990 年代にかけて、ヨーロッパの研究機関を中心に活発に研究されている。そのほとんどは、より良いオペラ歌唱の合成や、それを通じたオペラ歌唱の生成メカニズムの解明を目的とした、いわば西洋文化におけるオペラ歌唱の重要性を反映した取り組みと言えるものであった。中でも、映画「カストラート (伊 *Farinelli Il Castrato*)」において IRCAM (フランス国立音響音楽研究所) のグループが作成した去勢された男性歌手 (カストラート) の歌声合成音 [14] や、KTH (スウェーデン王立工科大学) の Sundberg らによるホルマント合成型の歌声合成システム [15]、Cook らの声道音響管の物理モデルを用いた歌声合成システム [16] が代表的である。

近年では、話声の音声合成分野で確立された波形接続方式や HMM (隠れマルコフモデル) 合成方式といったコーパスベースの合成手法を採用することで、実用性の高い歌声合成システムが提案されてきている [17-19]。中でも YAMAHA による VOCALOID [18] の技術を用いた、クリプトン・フューチャー・メディア社の「初音ミク」や「鏡音リン・レン」は、市販の歌声合成ソフトウェアとしては異例の販売数を記録しており、エンドユーザが多様な歌声を自由に合成して楽しめる時代が到来したと言える。

ほかにも、スキャット歌唱の分析・加工・合成システム [20]、歌詞の朗読音声歌声に変換する歌声合成システム [21]、異なる歌唱スタイルを混ぜ合わせる歌声モーフィング [22]、更にはユーザの歌い方を真似ることが可能な歌声合成システム [23] が提案されており、歌声合成研究の多様化が進んでいる。

また、国際会議 Interspeech 2007 では、歌声合成のスペシャルセッションが企画され、世界中の 6 グループが提案している歌声合成システムが一同に介して、歌声合成音の品質を紹介し合った [24]。このことから、音声信号処理分野における歌声合成への注目が世界中で高まっていることが分かる。

以上のように、現在、歌声合成は、学術的及び産業応用的に特に注目されている歌声情報処理技術の一つである。今後は、歌唱スタイルや歌唱者の違いも柔軟に表現し、表情豊かな歌声が合成可能なシステムの登場が期待される。

4. 歌詞を聴く計算機

歌詞は歌声が表現する情報の中でも、最も重要な要素の一つであり、計算機による歌詞の理解を目指した研究が盛んになっている。

4.1 歌詞の認識

歌詞の認識とは、歌詞が未知の楽曲の音響信号から、その楽曲の歌詞を自動的に書き起こすことを目指す研究である。これは、音声認識の歌声版と捉えることができる。しかし、歌声を対象とした場合、1 章で述べた変動の大きさや伴奏音といった難しさがあるため、伴奏を含む歌声にも適用可能な歌詞認識はまだ実現されていない。しかし、歌詞の認識は、歌詞による楽曲検索など様々な応用が可能で、今後の研究の発展が期待される。

無伴奏の歌唱を対象にした研究例としては、単独歌唱の自動採譜システムに歌詞の認識を組み込んだ研究 [25]、歌声の音高と歌詞の認識精度との関係を調べた研究 [26]、歌声のような高い音高においてもロバストな音声のモデル化手法に関する研究 [27] や、ハミング検索 (6.1 節) において音高だけでなく歌詞も考慮することで性能向上に取り組んだ研究 [28] がある。一方、伴奏を含む歌唱を対象にした研究は、4.2 節の歌詞同期以外ではほとんどないが、音素境界が既知という限定された条件下で歌声の音素認識に取り組んだ研究 [29] はある。ほかにも、歌詞の認識ではないが、伴奏付き歌唱中の歌詞の言語同定に取り組んだ研究 [30] がある。

4.2 歌詞の同期

歌詞の同期とは、歌詞が既知の楽曲に対して、歌詞と音楽音響信号とを時間的に対応付ける研究である。この研究課題は、音声認識の研究で使われる HMM に基づく強制アラインメント技術が対象とする問題に近いが、前節でも述べた歌声特有の難しさがある。しかし、歌詞の認識に比べると、問題設定を単純化したことで実用的な性能が得られやすく、一部の例外 [31] を除いて、伴奏付きの歌唱を対象にしている。そうした伴奏付きの歌唱を対象とする研究は、強制アラインメントを使用しない研究 [32-34] と、強制アラインメント使用する研究 [35-38] の二つに分類できる。

強制アラインメントを用いない場合、歌声の音韻的な特徴は利用せず、手がかりとして、声調言

語である広東語特有の性質 [32] や音韻の持続時間長 [34], 楽曲の標準 MIDI ファイル [33] などを用いる。ただし, 適用できる範囲に制約は大きい。

一方, 強制アラインメントを用いる場合, 伴奏の影響をどのように低減させるかが重要となる。そのために, 歌声の F_0 推定をして歌声を伴奏音から分離する手法 [39–42] や, 歌声が存在する区間を検出する手法 (歌声区間検出手法) [43] などと併用することが多い。そうした歌詞同期研究の具体例としては, 歌声区間検出手法を用いた研究 [35] や, 文献 [39] の手法で歌声を分離した後に歌声区間検出をする研究 [36], 更に調波構造で分離できない無声子音を別に検出する研究 [38] などがある。また, 別のアプローチとして, 強制アラインメントの探索範囲に音楽的知識による制約を設けることで, 性能向上を図った研究 [37] がある。

5. 声色が分かる計算機

歌声の声色 (声質) は, 誰が歌っているかという情報を与えるだけでなく, 声を聴いたときの人間の聴感的印象を大きく左右する大切な要素である。

5.1 歌手名の同定

歌手名の同定とは, あらかじめ DB に登録された歌手の中から, 入力した楽曲の歌手が誰かを特定する研究で, 2000 年代初めから多くの研究例がある [44–54]。一般的には, ケプストラム法や線形予測法 (LPC) などに基づく特徴量を抽出後, 混合ガウス分布 (GMM) や HMM, サポートベクターマシン (SVM) などを用いたパターン認識手法により歌手名を同定する。具体例としては, 歌声区間検出手法によって歌声と判断された区間のみを同定に使用する研究 [45–47, 49, 53], 歌声区間検出に加えて伴奏音の影響を低減させる処理をした研究 [48, 50, 51, 54], ヴィブラートの個人性を利用した研究 [52] などがある。また, 問題設定は多少異なるが, ある歌手の歌声が楽曲に含まれているかを 2 値判定する研究もある [55]。

5.2 歌声の類似度

楽曲間で歌手の歌声 (声色) がどれだけ似ているかを, 歌声の類似度として定量的に求める手法は応用範囲が広く, 歌声の類似度に基づいた楽曲の自動分類 [56] や, ある曲の歌手の歌声とよく似た歌声を持つ楽曲を検索する検索システム [57] など, 様々な研究事例がある。いずれも歌手名同定

のための手法を発展させた研究であり, 静的なスペクトルの類似度に基づいて歌声が似ているかどうかを推定する。今後, 歌い直しなどの歌声の動的な要素も考慮に入れるなど, 研究の更なる発展が期待される。

6. 歌声を検索する計算機

近年, 個人が大量の楽曲にアクセスできるようになり, 様々な方法で聴きたい楽曲を効率良く検索・ブラウジングする音楽情報検索 (MIR: Music Information Retrieval) の重要性が高まっている。本章では, そうした検索方法の中から, 歌唱を対象としたハミング検索, ヴォーカルの声質に基づいた歌声検索, 口ドラム検索について紹介する。

6.1 ハミング検索

ハミングとは「ラララー」等の歌詞を伴わない歌唱を指し, ハミング検索 (QBH: Query-by-Humming) は, それを検索キーとして楽曲を検索する方法である。書誌情報 (タイトルやアーティスト名) 以外を利用する「コンテンツ内容に基づく音楽情報検索」の先駆けであり, 1990 年代の黎明期の研究 [58–60] を経て, 数多くの研究がなされてきた (詳しい紹介は, 文献 [1, 61, 62] に譲る)。最近では, QBSH (Query-by-Singing/Humming) とも呼ばれ, 音楽情報検索のコンテストである MIREX (Music Information Retrieval Evaluation eXchange) [63] の課題として 2006 年~2008 年に採用されている。また, うたごえ検索 [64] や midomi [65] といった, ハミング検索を利用した音楽検索サービスも登場している。

ほかにも, F_0 軌跡の動的変動を相平面 ($F_0-\Delta F_0$ 平面) で表現する方法 [66] や, 歌声から F_0 推定せずに音程を確率モデルで表す方法 [67], ユーザ歌唱を学習して検索性能を上げる方法 [68] なども提案されている。また, F_0 以外の特徴量も検索で利用する試みとして, 歌詞情報を検索に利用する方法 [28] や, MFCC (Mel-Frequency Cepstral Coefficients) やホルマント周波数等を利用する方法 [69] などもある。

6.2 歌声検索

5.2 節で述べた歌声の類似度を利用して, 歌手の声質に基づいて音楽を検索するシステム Vocal-Finder [57] が提案されている。このシステムでは, ユーザが検索キーとして自分好みの声色の歌唱を

含む楽曲を入力すると、楽曲 DB の中から、その声色に似た歌唱の楽曲を探しだしてユーザに提示する。これにより、ユーザは自分の好みの歌声を持つ今まで知らなかった楽曲を発見することができる。

6.3 ロドラム検索

口(くち)ドラムとは、ドラム音を「ドンタンドタン」のように擬音語で表現したり、音響的に模倣して発声したりした歌唱音声を指す。海外では主に後者を扱い、それを Beatboxing と呼ぶことが多い。口ドラム検索 (Query-by-Beatboxing) は、そのような歌唱音声を認識し、対応するドラム音もしくはドラムパターン (1 小節を最小単位とするドラムスの演奏パターン) を得る方法である。

擬音語表現に近い発声を認識するには、主に音声認識と同様の枠組みで、HMM を用いて単音(「ドン」など)ごとにモデル化する方法 [70, 71]、音素 (/d/, /o/, /N/ など) ごとにモデル化する方法 [72] がある。一方、Beatboxing (ドラム音を音響的に模倣した音声) を対象とした研究では、主にドラム音認識で利用されている音響特徴量に基づき、SVM やニューラルネットワークなどの機械学習を用いて認識している [73-75]。

7. 歌声を評価する計算機

歌のうまさ(歌唱力)は、歌唱を特徴付ける重要な要素の一つである。歌唱力を決定付ける音響的な要素の解明は、歌唱指導や訓練場面に有用なだけでなく、音楽情報検索や歌声合成などへの応用も期待できる。本章では、歌唱力に関係する音響特徴と自動評価、アプリケーションを紹介する。

7.1 歌唱力に関係する音響特徴

ヴィブラートと singer's formant を有することが典型的な good voice (良い歌声) の条件であり [76]、それらが「歌声らしい」という聴感的印象に影響を与えているという知見がある [7]。また、2.1 節のオーバシュートは、歌唱力の差によらず生起するが、プロによる歌唱の方が、アマチュアよりも変動が小さいという知見がある [77]。

逆に vocal ugliness (歌声の不快さ) の条件として、不規則なヴィブラートや音高の不一致、音源スペクトルにおける乏しい基本波振幅、及び singer's formant の欠如が指摘されている [15]。

また、歌唱中のブレス(息継ぎ、吸気)につい

て、その仕方や発声位置は歌唱力と関連する可能性がある。まだその関連性の詳細は調査されていないが、ブレスの音響分析や自動検出は研究されている [78, 79]。

7.2 歌唱力の自動評価

歌唱力を自動的に評価するシステムとしては、カラオケの採点機能が普及しており、そこでは主に評価用の楽譜情報(音高)からの差異に基づいて採点している。最近では、ヴィブラート検出、しゃくり(ポルタメント)検出などの機能も備えている。また、音高や音長の一致に基づいて自動評価する研究がある [80]。

一方で、楽譜を用いずに歌唱力を自動的に評価する研究もある。そのような研究としては、相対音高とヴィブラートに関する音響特徴を用いた SVM による方法 [81] や、MFCC やメルフィルタバンク係数を特徴量とした GMM による方法 [82] がある。後者では、singer's formant に相当する周波数帯域が評価に有効であった。

7.3 歌唱力に関連したアプリケーション

歌唱力に関連したアプリケーションとして、カラオケの自動採点機能以外に、歌唱力向上(歌唱トレーニング)の支援が考えられる。具体例として、ユーザ歌唱の分析結果をリアルタイムに可視化してフィードバックする方法が提案されている [83-86]。特に、目標とする音高に対して、ユーザ歌唱の音高がどれくらい外れているかが分かるように表示することで、ユーザ歌唱の音高の正確さが向上することが報告されている [83, 84]。

8. 歌声情報処理の今後

ここまで歌声に関する研究が注目を集めた時代は過去にたく、「歌声情報処理」の研究は、今後も急速に進展していくことが予想される。本解説で述べた以外にも様々な研究課題が未解決のまま残っており、また、年々新たな課題も提案され、学会発表されている。学際的な研究分野であるため、今回はそのすべての側面は紹介できなかったが、心理学 [87]、生理学 [88]、声楽 [89] 等の歌声を取り巻く様々な知見も、信号処理、機械学習、インタフェース等と合わせて考慮していくことが、今後大切になると考えられる。

最初にも述べたように、歌声は、音声と音楽の両方の側面を持つ。現在、音声言語情報処理と音

楽情報処理の研究分野は、お互いに影響を与えつつもまだ接点は多くない。しかし、将来的には、音声と音楽を別々に考えずに、それらの総合的な情報処理の実現を目指す「音情報処理」という分野を確立していきたいと我々は考えている。歌声情報処理の研究は、正にそのための王道的アプローチの一つであり、成功の鍵を握っている。同分野が、より多くの人達の興味を集め、今後も更に発展していくことを期待したい。

謝 辞

3章の執筆にあたり、北海道医療大学 榊原健一氏から有益なコメントをいただいた。

文 献

- [1] 後藤真孝, 平田圭二, “音楽情報処理の最近の研究,” 音響学会誌, 60, 675–681 (2004).
- [2] 片寄晴弘, 後藤真孝, “音楽のデザイン転写技術の開発にむけて—CrestMuse プロジェクトの「価値」創出視点からの紹介—,” 人工知能学会第20回全国大会 (2006).
- [3] 後藤真孝, “研究会千夜一夜, 音楽情報処理の研究を始めませんか?—音楽情報科学研究会 (SIGMUS)—,” 情報処理, 48, 1046–1047 (2007).
- [4] T. Saitou, M. Unoki and M. Akagi, “Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis,” *Speech Commun.*, 5, 267–277 (2005).
- [5] J. Sundberg, *The Science of Singing Voice* (Northern Illinois University Press, DeKalb, 1987).
- [6] I. Nakayama, “Comparative studies on vocal expression in Japanese traditional and Western classical-style singing, using a common verse,” *Proc. ICA 2004*, pp.1295–1296 (2004).
- [7] 齋藤 毅, 辻 直也, 鶴木祐史, 赤木正人, “歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析,” 音響学会誌, 46, 405–417 (2008).
- [8] 中山一郎, “日本語を歌・唄・謡う,” 音響学会誌, 11, 688–693 (2003).
- [9] M. Lesaffre, K. Tanghe, G. Martens, D. Moelants, M. Lemanand, B.D. Baets, H.D. Meyer and J.-P. Martens, “The MAMI query-by-voice experiment, Collecting and annotating vocal queries for music information retrieval,” *Proc. ISMIR 2003*, pp.65–71 (2003).
- [10] E. Unal, S.S. Narayanan, H.-H. Shih, E. Chew and C.-C.J. Kuo, “Creating data resources for designing usercentric frontends for query-by-humming systems,” *Multimedia Syst.*, 10, 475–483 (2005).
- [11] J.-S.R. Jang, “QBSH, A corpus for designing QBSH (Query by Singing/Humming) systems,” <http://www.cs.nthu.edu.tw/~jang/>.
- [12] 後藤真孝, 橋口博樹, 西村拓一, 岡 隆一, “RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース,” 情報処理学会論文誌, 45, 728–738 (2004).
- [13] 後藤真孝, 西村拓一, “AIST ハミングデータベース: 歌声研究用データベース,” 情報処理学会研究報告, 2005-MUS-61, pp. 7–12 (2005).
- [14] P. Depalle, G. Garcia and X. Rodet, “A virtual castrato,” *Proc. ICMC '94*, pp.357–360 (1994).
- [15] J. Sundberg, “The KTH synthesis of singing,” *Adv. Cognit. Psychol. (Special issue on Music Performance)*, 2, 131–143 (2006).
- [16] P.R. Cook, “Identification of control parameters in an articulatory vocal tract model, with applications to the synthesis of singing,” *PhD Thesis, Stanford Univ.* (1991).
- [17] J. Bonada and X. Serra, “Synthesis of the singing voice by performance sampling and spectral models,” *IEEE Sig. Process. Mag.*, 24, 67–79 (2007).
- [18] 剣持秀紀, 大下隼人, “歌声合成システム VOCALOID,” 情報処理学会研究報告, 2007-MUS-072, pp.25–28 (2007).
- [19] 酒向慎司, 宮島千代美, 徳田恵一, 北村 正, “隠れマルコフモデルに基づいた歌声合成システム,” 情報処理学会論文誌, 45, 719–727 (2004).
- [20] 河原英紀, 片寄晴弘, “高品質音声分析変換合成システム STRAIGHT を用いたスキヤット生成研究の提案,” 情報処理学会論文誌, 43, 208–218 (2002).
- [21] 齋藤 毅, 後藤真孝, 鶴木祐史, 赤木正人, “SingBy-Speaking: 歌声知覚に重要な音響特徴を制御して話声を歌声に変換するシステム,” 情報処理学会研究報告, 2008-MUS-74-5, pp.25–32 (2008).
- [22] 河原英紀, 生駒太一, 森勢将雅, 高橋 徹, 豊田健一, 片寄晴弘, “モーフィングに基づく歌唱デザインインタフェースの提案と初期検討,” 情報処理学会論文誌, 48, 3637–3648 (2007).
- [23] 中野倫靖, 後藤真孝, “VocaListener: ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案,” 情報処理学会研究報告, 2008-MUS-75, pp.49–56 (2008).
- [24] Interspeech 2007, Synthesis of Singing Challenge, http://www.interspeech2007.org/Technical/synthesis_of_singing_challenge.php.
- [25] C.-K. Wang, R.-Y. Lyu and Y.-C. Chiang, “An automatic singing transcription system with multilingual singing lyric recognizer and robust melody tracker,” *Proc. Eurospeech 2003*, pp.1197–1200 (2003).
- [26] 尾関弘尚, 鎌田貴幸, 後藤真孝, 速水 悟, “歌声の歌詞認識における音高の影響について,” 音講論集, pp.637–638 (2003.10).
- [27] A. Sasou, M. Goto, S. Hayamizu and K. Tanaka, “An auto-regressive, non-stationary excited signal parameter estimation method and an evaluation of a singing-voice recognition,” *Proc. ICASSP 2005*, pp.1-237–240 (2005).
- [28] M. Suzuki, T. Hosoya, A. Ito and S. Makino, “Music information retrieval from a singing voice using lyrics and melody information,” *EURASIP J. Adv. Signal Process.*, 2007 (2007).
- [29] M. Gruhne, K. Schmidt and C. Dittmar, “Phoneme recognition in popular music,” *Proc. ISMIR 2007*, pp.369–370 (2007).
- [30] W.-H. Tsai and H.-M. Wang, “Automatic identification of the sung language in popular music recordings,” *J. New Music Res.*, 36, 105–114 (2007).
- [31] A. Loscos, P. Cano and J. Bonada, “Low-delay singing voice alignment to text,” *Proc. ICMC 99* (1999).
- [32] C.H. Wong, W.M. Szeto and K.H. Wong, “Automatic lyrics alignment for Cantonese popular music,” *Multimedia Syst.*, 4-5, 307–323 (2007).
- [33] M. Müller, F. Kurth, D. Damm, C. Fremerey and M. Clausen, “Lyrics-based audio retrieval and multimodal navigation in music collections,” *Proc. ECDDL 2007*, pp.112–123 (2007).
- [34] M.-Y. Kan, Y. Wang, D. Iskandar, T.L. Nwe and

- A. Shenoy, "LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals," *IEEE Trans. Audio Speech Lang. Process.*, 16, 338–349 (2008).
- [35] K. Chen, S. Gao, Y. Zhu and Q. Sun, "Popular song and lyrics synchronization and its application to music information retrieval," *Proc. MMCN '06* (2006).
- [36] H. Fujihara, M. Goto, O. Jun, K. Komatani, T. Ogata and H.G. Okuno, "Automatic synchronization between lyrics and music CD recordings based on Viterbi alignment of segregated vocal signals," *Proc. ISM 2006*, pp. 257–264 (2006).
- [37] D. Iskandar, Y. Wang, M.-Y. Kan and H. Li, "Syllabic level automatic synchronization of music signals and text lyrics," *Proc. ACM Multimedia 2006*, pp. 659–662 (2006).
- [38] H. Fujihara and M. Goto, "Three techniques for improving automatic synchronization between music and lyrics, fricative sound detection, filler model, and novel feature vectors for vocal activity detection," *Proc. ICASSP 2008*, pp. 69–72 (2008).
- [39] M. Goto, "A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals," *Speech. Commun.*, 43, 311–329 (2004).
- [40] M. Ryynänen and A. Klapuri, "Transcription of the singing melody in polyphonic music," *Proc. ISMIR 2006*, pp. 222–227 (2006).
- [41] J.-L. Durrieu, G. Richard and B. David, "Singer melody extraction in polyphonic signals using source separation methods," *Proc. ICASSP 2008*, pp. 169–172 (2008).
- [42] 藤原弘将, 後藤真孝, 奥乃 博, "歌声の統計的モデル化とビタビ探索を用いた多重奏中のボーカルパートに対する音高推定手法," *情報処理学会論文誌*, 49 (2008).
- [43] T.L. Nwe and Y. Wang, "Automatic detection of vocal segments in popular songs," *Proc. ISMIR 2004*, pp. 138–145 (2004).
- [44] B. Whitman, G. Flake and S. Lawrence, "Artist detection in music with minnowmatch," *Proc. NNSP 2001*, pp. 559–568 (2001).
- [45] A.L. Berenzweig, D.P. W. Ellis and S. Lawrence, "Using voice segments to improve artist classification of music," *Proc. AES-22 Int. Conf. Virtual, Synthetic and Entertainment Audio* (2002).
- [46] Y.E. Kim and B. Whitman, "Singer identification in popular music recordings using voice coding features," *Proc. ISMIR 2002*, pp. 164–169 (2002).
- [47] T. Zhang, "Automatic singer identification," *Proc. ICME 2003*, Vol. I, pp. 33–36 (2003).
- [48] M.A. Bartsch, "Automatic singer identification in polyphonic music," *PhD Thesis, Univ. Michigan* (2004).
- [49] N.C. Maddage, C. Xu and Y. Wang, "Singer identification based on vocal and instrumental models," *Proc. ICPR '04*, Vol. 2, pp. 375–378 (2004).
- [50] 藤原弘将, 北原鉄朗, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃 博, "伴奏音抑制と高信頼度フレーム選択に基づく楽曲の歌手名同定手法," *情報処理学会論文誌*, 47, 1831–1843 (2006).
- [51] W.-H. Tsai and H.-M. Wang, "Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals," *IEEE Trans. Audio Speech Lang. Process.*, 14, 330–341 (2006).
- [52] T.L. Nwe and H. Li, "Exploring vibrato-motivated acoustic features for singer identification," *IEEE Trans. Audio Speech Lang. Process.*, 15, 519–530 (2007).
- [53] J. Shen, B. Cui, J. Shepherd and K.-L. Tan, "Towards efficient automated singer identification in large music databases," *Proc. SIGIR '06*, pp. 59–66 (2006).
- [54] H. Fujihara, T. Kitahara, M. Goto, K. Komatani, T. Ogata and H.G. Okuno, "Singer identification based on accompaniment sound reduction and reliable frame selection," *Proc. ISMIR 2005*, pp. 329–336 (2005).
- [55] W.-H. Tsai and H.-M. Wang, "Automatic detection and tracking of target singer in multi-singer music recordings," *Proc. ICASSP 2004*, pp. 221–224 (2004).
- [56] W.-H. Tsai, H.-M. Wang, D. Rodgers, S.-S. Cheng and H.-M. Yu, "Blind clustering of popular music recordings based on singer voice characteristics," *Proc. ISMIR 2003*, pp. 167–173 (2003).
- [57] H. Fujihara and M. Goto, "A music information retrieval system based on singing voice timbre," *Proc. ISMIR 2007*, pp. 467–470 (2007).
- [58] T. Kageyama, K. Mochizuki and Y. Takashima, "Melody retrieval with humming," *Proc. ICMC 1993*, pp. 349–351 (1993).
- [59] A. Ghias, J. Logan, D. Chamberlin and B. Smith, "Query by humming: Musical information retrieval in an audio database," *Proc. ACM Multimedia 1995*, Vol. 95, pp. 231–236 (1995).
- [60] T. Sonoda, M. Goto and Y. Muraoka, "A WWW-based melody retrieval system," *Proc. ICMC 1998*, pp. 349–352 (1998).
- [61] R. Dannenberg, W. Birmingham, B. Pardo, C. Meek, N. Hu and G. Tzanetakis, "A comparative evaluation of search techniques for query-by-humming using the MUSART testbed," *J. Am. Soc. Inf. Sci. Technol.*, 58, 687–701 (2007).
- [62] E. Unal, E. Chew, P. Georgiou and S. Narayanan, "Challenging uncertainty in query by humming systems: A fingerprinting approach," *IEEE Trans. Audio Speech Lang. Process.*, 16, 359–371 (2008).
- [63] J.S. Downie, "The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research," *Acoust. Sci. & Tech.*, 29, 247–255 (2008).
- [64] ウタゴエ株式会社: うたごえ検索, <http://www.utagoie.com/jp/service/search.html>.
- [65] Melodis Corporation: midomi, <http://www.midomi.co.jp/>.
- [66] Y. Ohishi, M. Goto, K. Ito and K. Takeda, "A stochastic representation of the dynamics of sung melody," *Proc. ISMIR 2007*, pp. 371–372 (2007).
- [67] 市川拓人, 鈴木基之, 伊藤彰則, 牧野正三, "複数の音程特徴量によるハミング入力楽曲検索システムの高精度化," *情報処理学会研究報告*, 2008-MUS-74, pp. 7–12 (2008).
- [68] D. Little, D. Raffensperger and B. Pardo, "A query by humming system that learns from experience," *Proc. ISMIR 2007*, pp. 335–338 (2007).
- [69] A. Duda, A. Nurnberger and S. Stober, "Towards query by singing / humming on audio databases," *Proc. ISMIR 2007*, pp. 331–334 (2007).
- [70] O. Gillet and G. Richard, "Drum loops retrieval from spoken queries," *J. Intell. Inf. Syst.*, 24, 159–177 (2005).
- [71] O. Gillet and G. Richard, "Indexing and querying drum loops databases," *Proc. CBMI 2005* (2005).
- [72] 中野倫靖, 緒方 淳, 後藤真孝, 平賀 譲, "口ドラム

- 認識手法とそのドラム譜入力システムへの応用,” 情報処理学会論文誌, 48, 386–397 (2007).
- [73] A. Kapur, M. Benning and G. Tzanetakis, “Query-by-beat-boxing: Music retrieval for the DJ,” *Proc. ISMIR 2004*, pp. 170–177 (2004).
- [74] A. Hazan, “Towards automatic transcription of expressive oral percussive performances,” *Proc. IUI 2005*, pp. 296–298 (2005).
- [75] E. Sinyor, C. McKay, R. Fiebrink, D. McEnnis and I. Fujinaga, “Beatbox classification using ACE,” *Proc. ISMIR 2005*, pp. 672–675 (2005).
- [76] W.T. Bartholomew, “A physical definition of “good voice-quality” in the male voice,” *J. Acoust. Soc. Am.*, 55, 838–844 (1934).
- [77] 齋藤 毅, 鷗木祐史, 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).
- [78] D. Ruinskiy and Y. Lavner, “An effective algorithm for automatic detection and exact demarcation of breath sounds in speech and song signals,” *IEEE Trans. Audio Speech Lang. Process.*, 15, 838–850 (2007).
- [79] T. Nakano, J. Ogata, M. Goto and Y. Hiraga, “Analysis and automatic detection of breath sounds in unaccompanied singing voice,” *Proc. ICMPC 10* (2008).
- [80] P. Lal, “A comparison of singing evaluation algorithms,” *Proc. Interspeech 2006*, pp. 2298–2301 (2006).
- [81] 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, 48, 227–236 (2007).
- [82] P. Prasert, K. Iwano and S. Furui, “An automatic singing voice evaluation method for voice training systems,” 音講論集, pp. 911–912 (2008.3).
- [83] D.M. Howard and G.F. Welch, “Microcomputer-based singing ability assessment and development,” *Appl. Acoust.*, 27, 89–102 (1989).
- [84] 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,” 信学論, J84-D-II, 1933–1941 (2001).
- [85] D. Hoppe, M. Sadakata and P. Desain, “Development of real-time visual feedback assistance in singing training: A review,” *J. Comput. Assist. Learn.*, 22, 308–316 (2006).
- [86] T. Nakano, M. Goto and Y. Hiraga, “Mirusinger: A singing skill visualization interface using real-time feedback and music CD recordings as referential data,” *Proc. ISM 2007 Workshops*, pp. 75–76 (2008).
- [87] D. Deutsch, Ed., *The Psychology of Music* (Academic Press, New York, 1982).
- [88] I.R. Titze, *Principles of Voice Production* (The

National Center for Voice and Speech, Denver, 2000).

[89] フレデリック フースラー, イヴォンヌ ロッド・マーリング, うたうこと 発声器官の肉体的特質—歌声のひみつを解くかぎ (音楽之友社, 東京, 1987).



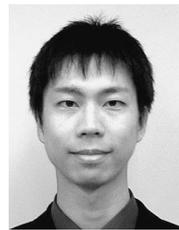
後藤 真孝

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主任研究員。筑波大学大学院准教授(連携大学院), 統計数理研究所客員准教授を兼任。科学技術分野の文部科学大臣表彰 若手科学者賞, ドコモ・モバイル・サイエンス賞 基礎科学部門 優秀賞等 22件受賞。日本音響学会, 情報処理学会, 電子情報通信学会, 日本音楽知覚認知学会各会員。



齋藤 毅

2000年金沢大・工・電気情報工卒, 2002年北陸先端科学技術大学院大学情報科学研究科博士前期課程修了。2006年同博士後期課程修了。博士(情報科学)。同年ATR認知情報科学研究所, 2007年から産業技術総合研究所特別研究員。歌声知覚及び合成に関する研究に従事。2006年聴覚研究会研究奨励賞受賞。日本音響学会, 情報処理学会, ISCA 各会員。



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(情報科学)。現在, 産業技術総合研究所特別研究員。音楽デザイン転写・音響信号理解に基づく音インタフェースに従事。日本音楽知覚認知学会 2006年度春季研究発表会研究選奨, インタラクシオン2007インタラクティブ発表賞受賞。日本音響学会, 情報処理学会, 日本音楽知覚認知学会, 各会員。



藤原 弘将

2005年京都大学工学部情報学科卒業。2007年同大学大学院情報学研究科知能情報学専攻修士課程修了。同年, 産業技術総合研究所に入所し, 現在に至る。音楽情報処理, 音楽情報検索, 音声情報処理に興味を持つ。平成19年度山下記念研究賞受賞。日本音響学会, 情報処理学会, 電子情報通信学会各会員。