

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 4 月 19 日現在

機関番号：13903

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700191

研究課題名（和文） 多彩な歌唱表現を自動学習する制御性に優れた歌唱音声合成システムの研究

研究課題名（英文） A study on trainable singing voice synthesis system capable of representing personal characteristics and singing styles

研究代表者

酒向 慎司（SAKO SHINJI）

名古屋工業大学・大学院工学研究科・助教

研究者番号：30396791

研究成果の概要（和文）：

本研究では、これまでに提案してきた隠れマルコフモデル（HMM）に基づいた歌唱合成手法の枠組みを用いて、多様な声質や歌唱表現、多言語歌唱を実現する手法について研究を行った。また、新たに大規模な歌唱データベースを構築し、新たな歌唱モデルを作成する際のガイドラインを検討した。

研究成果の概要（英文）：

In this research, we developed singing voice, singing style and singing voice for other language by using HMM-based singing voice synthesis framework that we have proposed. On the other hand, we also developed a large scale singing voice database and designed a guide line to make new singing voice model.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	1,200,000	360,000	1,560,000
2010 年度	700,000	210,000	910,000
2011 年度	700,000	210,000	910,000
総計	2,600,000	780,000	3,380,000

研究分野：総合領域

科研費の分科・細目：情報学、知覚情報処理・知能ロボティクス

キーワード：音声情報処理、歌唱音声合成

1. 研究開始当初の背景

計算機によって歌唱を合成させる試みは古くから行われてきた。最近では Vocaloid などに代表される、歌詞とメロディから人間の歌声を合成する商用ソフトウェアが大きな成果を収めているが、歌声らしい合成音を自由に・自在に作ることはまだ難しい。一般的なテキスト音声合成では、読み上げ音声の合成という基準では一定のレベルに達したものの、意思や感情表出など“人間らしさ”の欠如という問題が残されており、これと同形の問題ととらえる。

本研究では、歌唱における表情を、曲や詩

を歌唱者が解釈することによって生じる変形として考える。つまり、これらの変形が曲や詩という制約によって確率的に揺らいで観測されると仮定することで、その振る舞いを確率的な生成モデルとして学習することができ、これまでに提案してきた歌唱音声合成手法に組み込むことで、表情が付与された歌唱を合成することができると考えた。

一方で、先に挙げた歌唱音声合成ソフトウェアの成功要因の一つに、単に楽譜通りに歌唱するだけでなく、歌唱表現を編集できる余地があった点が大きいと考えている。これは、ある種の“作りこみ”のテクニックとされる

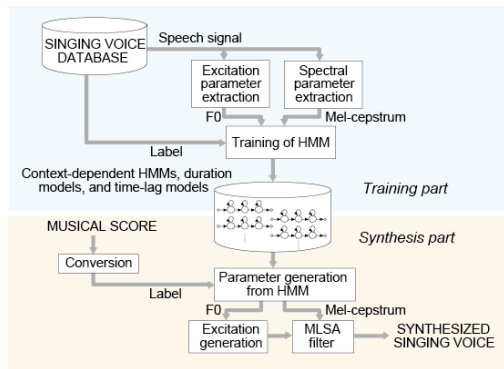


図1. 歌声合成 HMM の概要

ものであるが、コンテンツ作成の側面から考えると、このような制御性あるいは加工性が重要な要素であることが伺える。このような要求には、本研究のように歌唱音声の生成過程をパラメトリックなモデルとして扱うアプローチがより有効である。

これまでに提案してきた隠れマルコフモデル (Hidden Markov Models, HMM) に基づいた歌唱音声合成システムは、一般的な素片接続方式の合成方式とは本質的に異なり、歌唱音声における音の高さ、音色・大きさ、時間構造の変化を確率的な枠組みでモデル化したものである (図1)。このように音声の生成過程がモデル化されることにより、本研究の二つの主題である多様性はモデル変換によって、加工性・制御性はモデルパラメータの操作によって実現することができる。

2. 研究の目的

歌唱表現の多様性は、個人性、歌唱法、曲や詩の解釈に由来するものなどが考えられるが、主に曲や詩の解釈に着目する。解釈は曲や詩という明確なシンボルに対してどう振る舞うかという変形として考えると、曲や詩が歌唱音声の生成モデルにどのような影響を及ぼすかというモデル変形の問題へと帰着させることができる。例えば、音符と音素の列として対応する音声を生成するのではなく、メロディや詩を考慮した歌唱音声を合成することに相当する。このような歌唱表現システムを、実際の歌唱データから学習可能であることを示す。

歌唱音声の加工性・制御性における課題については、歌声合成モデルにおいて、どのようなパートが歌声の特徴に相当するかを調査し、実用上で有効な歌唱音声パラメータの簡便な操作インターフェースを実現し、歌唱合成アプリケーションへと組み込む。また、歌声に関する研究に有用な大規模な歌唱音声データベースを構築することで、その効率的な構築手法を確立することも目的とする。

統計モデルを用いた歌唱のモデル化手法

は、様々な応用が考えられ楽器演奏の個人性をモデル化するほか、その他時系列パタンの生成モデルの学習手法への適用を試みる。

3. 研究の方法

① 歌唱音声データの収録・整備

現状の歌唱音声データ (童謡 60 曲、男女各 1 名) ではデータの質と量の両面で不足するため、多様性のある歌唱データを整備する。音楽の知識と技量を備えたデータ提供者として、女性ボーカリストに依頼する。また、収録場所や機材確保の労力・コストを避けるため、音楽用のスタジオを利用する。収録内容は先行研究に準じるが、研究用データベースとしての取り扱いを考慮して、著作権の消滅した詞・曲を選択する。研究期間終了以降も研究用に公開して活用できるようにアルバイト謝金を使い、収録データを実用的なデータベースとして整備する。

② 歌唱の個人性と歌声変換の実験

歌唱の特徴は、主に音色 (声質)、ピッチパターン、時間構造によってモデル化される。歌唱合成はこれらの統計モデルによって生成され、歌手の特徴を再現することができる。複数の歌唱データごとにモデルを用意することは効率が悪いため、これらの歌唱合成モデルの一部を変形・置換することで歌唱の特徴を変化させ、異なる歌唱を合成することを試みる。その際、3つの統計モデルのうちどのパートが歌声の特徴にどの程度寄与しているかを調査する。

③ 多様な歌唱合成法の開発

HMM 歌唱合成では、スペクトルの概形、ピッチ、時間構造が音素単位でモデル化されている。言語間の音韻の類似性を用いることで、日本語歌唱の音響モデルから、別の言語の歌唱を合成する手法を開発する。また、分析合成系のパラメータを変化させることで、合成される歌唱の変形手法について検討する。

④ 歌唱合成デモシステムの作成

本研究の成果を自動作曲システム “Orpheus” の歌声合成エンジンへ組み込む。合成歌唱を生成する際に、単にメロディと歌詞を入力とするだけでなく、声質や、この自動作曲システムをより発展させることができると考えている。

⑤ 歌唱表現モデルの応用

HMM に基づいた歌唱合成モデルの考え方を応用し、楽器演奏における演奏表現のモデル化を試みる。これは、表情を備えた自動演奏の実現や、演奏者の個人性を識別・分析することに繋がる有用な技術となりうる。具体的

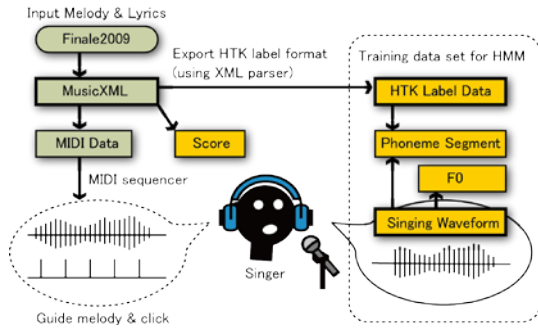


図 3. 歌声データ収録フロー

には、電子ピアノの演奏データである MIDI データを用い、演奏者によって生じる打鍵の強さやタイミングの違いを、楽譜の指示からどのぐらい逸脱するかという変動を確率モデルによって学習する。

また、音楽とは別の視点から、人間の動作パターン生成モデルを学習によって獲得する試みとして、手話の一部である指文字の合成手法に取り組む。これは、グローブ型のセンサから取得された指形状のデータ（各関節の曲げ角）の時間軌跡を、HMM によって学習・合成するものである。

4. 研究成果

①歌唱データベース構築と構築手法の確立

本稿では、HMM 歌唱合成での利用を念頭に、同一歌唱者による 1 時間程度の歌唱データの収録とデータベース化のフローを確立した。HMM 歌唱合成の学習データだけでなく、歌唱音声に関する様々な研究に寄与できるよう、表 1 に示すような内容の研究用データベースとして近く公開する予定である。

本データベースでは収録のための楽曲を音韻の分布を基準に選択したが、歌唱の特徴を広くカバーしているとは言えない。音高、テンポやリズムなどを考慮した選曲基準が望ましいが、今後の課題の一つとしたい。また、CrestMuse プロジェクトによる PEDB (Music Performance Expression Database) のような、演奏表情が付与された実演奏データベースが公開されている。本データベースでも、歌唱の表情という観点で、このような情報を付与することが有望である。楽譜に記載されている形式的なものだけでなく、ビブラートなど歌唱表現の特徴を XML のタグとして記述することで、幅広い活用が期待できる。

表 1. データベースの概要

曲数	60 曲
楽曲	童謡・唱歌
歌手	女性 (1 名)
録音条件	48kHz・24bit
収録場所	防音室

②歌唱の個人性と歌声変換の実験

歌声合成システムにおける歌唱の個人性がどのように知覚されているかを調べるため、歌声を構成する各特徴量が人間の知覚にどのような影響を与えるかを調査した。具体的には、HMM 歌唱合成の枠組みを用いて各特徴量を声質と歌い方の 2 つの要素に分類し、異なる 2 名の歌声モデルにおいて各々の声質と歌い方を組み合わせることで、部分的に 2 名の個人性を反映させた歌声を合成し、歌声を構成する各特徴量が個人性をどのように表現するか聴取実験により評価した。この実験の結果から、歌声における個人性は主に声質の違いに起因する要素が大きいことが確認できた。また、個人性を複数持った歌声を人間がどのように知覚するかという知見が得られた。

③多言語歌唱合成

図 2 に示されるような、音素決定木を用いて異なる言語間の発音様式の共通性を利用して、単一言語の歌唱データから学習された音響モデルから、他の言語の歌唱を合成するための手法を検討した。予備実験の結果から、若干の品質低下を許容することで、特定の音素レベルでモデルを区別するような言語依存の強いものは除外できることを示している。

合成実験の結果からは、とくに音符内の時間構造が不自然になる問題が見られた。その原因として、日本語歌唱と英語歌唱におけるリズムの性質の違いによるものと考えられる。特に学習用の歌唱データは童謡を中心としたものであり、その傾向が強く現れていると思われる。冒頭で述べたとおり、本来は複数の言語の歌唱データを用いることが望ましいが、音韻の再現性という視点だけでなく歌唱様式を補う意味でも必要である。複数の言語の歌唱データを収集し、本手法で提案したクラスタリングの木構造に言語間の共通性を持たせつつ、局所的には言語依存性を許容させることで、より汎用的な多言語歌唱手法が期待できる。

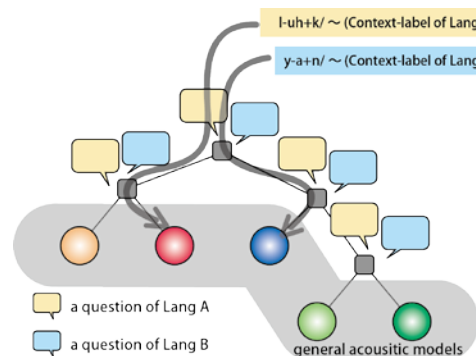


図 2. 決定木に基づく音韻変換

譜を考慮した演奏者の統計的モデル化手法—個人性と演奏傾向の比較分析—”, 日本音響学会 2011 年春季研究発表会, 3-1-19, pp.1081-1082, Mar. 11, 2011. (東京都市大学 世田谷キャンパス)

[4] 酒向 慎司, 北村 正, “隠れマルコフモデルに基づく指文字合成システムの開発”, 電子情報通信学会技術研究報告, Vol. 110, No. 384, WIT2010-88, pp.31-36. Jan. 22, 2011. (立命館大学 びわこ・草津キャンパス)

[5] Satoru Fukayama, Kei Nakatsuma, Shinji Sako, Takuya Nishimoto, Shigeki Sagayama, “Automatic Song Composition from the Lyrics exploiting Prosody of the Japanese Language”, Proc. of Sound and Music Computing Conference (SMC), pp.299-302, Jul. 22, 2010. (バルセロナ・査読あり)

[6] Satoru Fukayama, Kei Nakatsuma, Shinji Sako, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama, “Automatic Song Composition form Lyrics with Singing Voice Synthesizer”, Proc. of Intersinging, pp.1-4, Oct. 1, 2010. (東京大学、査読あり)

[7] 奥村 健太, 酒向 慎司, 北村 正, “楽譜を考慮した統計的演奏者モデル化手法と演奏者識別による評価”, 音響学会秋季研究発表会, 2-7-11, pp.895-896, Sep. 15, 2010. (関西大学・吹田キャンパス)

[8] 酒向 慎司, 北村 正, “HMM 歌唱合成のための歌唱データベースの構築”, 日本音響学会 2010 年秋季研究発表会, 3-P-30, pp.399-400, Sep. 16, 2010. (関西大学・吹田キャンパス)

[9] 奥村 健太, 酒向 慎司, 北村 正, “楽譜情報を用いたストリーム別コンテキストクラスタリングに基づく演奏者モデリング手法”, 音響学会春季研究発表会, 3-8-2, pp.999-1000, Mar. 10, 2010. (電気通信大学)

[10] 奥村 健太, 酒向 慎司, 北村 正, “楽譜情報のコンテキストクラスタリングに基づく演奏者のモデリング手法”, 情報処理学会音楽情報科学研究会, 2009-MUS-83, pp.1-6, Dec. 5, 2009. (国立音大)

[11] Satoru Fukayama, Kei Nakatsuma, Shinji Sako, Yuichiro Yonebayashi, T. H. Kim, Q. S. Wei, T. Nakano, Takuya Nishimoto, Shigeki Sagayama, “Orpheus: Automatic Composition System Considering Prosody of

Japanese Lyrics”, Proc. of International Conference on Entertainment Computing(ICEC), pp. 309-310, Sep. 4, 2009. (パリ、査読あり)

[12] Heiga Zen, Keiichiro Oura, Takashi Nose, Junichi Yamagishi, Shinji Sako, Tomoki Toda, Takashi Masuko, Alan W Black, Keiichi Tokuda, “Recent Development of the HMM-based Speech Synthesis System”, Proc. of Asia Pacific Signal and Information Processing Association (APSIPA), MP-SS1-2, Oct. 5, 2009. (札幌コンベンションセンター・査読あり)

[13] 酒向 慎司, 北村 正, “HMM 歌唱音声合成における日本語音響モデルを用いた英語歌唱合成の検討”, 日本音響学会秋季研究発表会, 2-2-10, pp.307-308, Sep. 16, 2009. (日本大学・郡山キャンパス)

[14] 奥村 健太, 酒向 慎司, 北村 正, “楽譜情報のコンテキストクラスタリングに基づく実演奏データからの演奏者モデリング手法の検討”, 音響学会秋季研究発表会, 3-5-2, pp.871-872, Sep. 17, 2009. (日本大学・郡山キャンパス)

[15] 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”, 情報処理学会音楽情報処理研究会, Vol.2009-MUS-80, No.5, pp.1-6, Mar. 21, 2009. (筑波大学・春日キャンパス)

[図書] (計 0 件)

[産業財産権]
○出願状況 (計 0 件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

○取得状況 (計 0 件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

[その他]

ホームページ等

<http://www.mmsp.nitech.ac.jp/~sako/>

<http://ngs.hil.t.u-tokyo.ac.jp/~orpheus/cgi-bin/>

6. 研究組織

(1) 研究代表者

酒向 慎司 (SAKO SHINJI)

名古屋工業大学・大学院工学研究科・助教

研究者番号：30396791