様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月　 １ 日現在

機関番号：１２６０８
研究種目：挑戦的萌芽研究
研究期間：2013～2014
課題番号：２５５４００６５

研究課題名（和文）ガウス過程回帰モデルに基づくノンパラメトリック音声合成の研究

研究課題名（英文）Research on speech synthesis using non-parametric modeling based on Gaussian process regression

研究代表者

小林　隆夫（Kobayashi, Takao）

東京工業大学・総合理工学研究科（研究院）・教授

研究者番号：７０１５３６１６

交付決定額（研究期間全体）：（直接経費）　　2,900,000円

研究成果の概要（和文）：隠れマルコフモデルに基づく音声合成手法の性能の限界を超えて，より多様で自然な合成音声を生成するために，ノンパラメトリックモデルを用いた音声合成手法を確立することをめざして研究を行った。提案する音声合成手法におけるモデル化はガウス過程回帰に基づいており，入力テキストに対してフレーム単位で音声合成に必要なパラメータを予測するためのカーネル関数の設計，計算量削減手法，ハイパーパラメータの自動最適化手法，ガウス過程分類を導入した韻律情報のモデル化手法などの検討を行った。

研究成果の概要（英文）：The purpose of the research is to develop a framework using non-parametric modeling for synthesizing more natural-sounding speech than the conventional HMM-based statistical parametric speech synthesis framework. The proposed modeling approach is based on Gaussian process regression (GPR) and GPR model is designed for directly predicting frame-level acoustic features from corresponding input linguistic information. We have proposed kernel functions for GPR-based speech synthesis and examined several techniques for computational cost reduction, hyper-parameter optimization, and prosody modeling using Gaussian process classification and GPR.

研究分野：音声情報処理

キーワード：テキスト音声合成　統計的パラメトリック音声合成　HMM音声合成　ガウス過程回帰　カーネル関数　フレームコンテキスト

１．研究開始当初の背景

任意の文章から音声を生成する「テキスト音声合成技術」において，統計的パラメトリックモデルに基づく手法がその柔軟性やコストの面で近年大きな注目を集めている。その中でも隠れマルコフモデルに基づく手法（HMM 音声合成）は最も代表的であり，最近では世界中の様々な言語を対象として幅広く研究や実用化が行われている。HMM 音声合成の研究が進み合成音声品質の向上する一方で，隠れマルコフモデルの構造自体に起因する根本的な性能の限界も指摘されるようになった。例えば，モデル化のための学習データを増やしても，波形選択・接続方式とは異なり合成音声の品質の向上が飽和してしまう，あるいは状態単位の区分定常性の仮定に基づいた平均化処理による過剰平滑化の影響で合成音声の自然性が劣化する，といった問題点である。

このためここ数年来，HMM 音声合成の枠組みにおける統計的パラメトリック音声合成の特長をいかしつつ，隠れマルコフモデルに代わる新たなモデルを導入する試みが盛んに検討されている。そして，大きな潮流となり始めているのが，ディープラーニング（深層学習）の枠組みに基づいたディープニューラルネット（DNN）やリカレントニューラルネット（RNN）を用いたモデル化・パラメータ生成手法である。

一方，本研究代表者らも，隠れマルコフモデルに代わる新たなモデルとして，ガウス過程回帰に基づくノンパラメトリックモデルを利用することを着想した。そして，音素単位レベルではあるが，提案手法に基づいた音声のスペクトルパラメータのモデル化とパラメータ生成の基礎的検討を始めた。

２．研究の目的

上述のような研究背景の下，本研究では，これまでの HMM 音声合成の性能の限界を超えて，より多様で自然な合成音声を生成するために，ノンパラメトリックモデルを用いる音声合成手法を確立することをめざす。

具体的には，HMM によるモデル化の限界が離散的な状態間の遷移に基づいた音声特徴量空間の表現に起因することを考慮し，これとは異なる枠組みであるガウス過程回帰に基づく音声の時系列データのノンパラメトリックなモデル化手法を検討する。そして，提案した音声の音響特徴量のノンパラメトリックモデル化手法をテキスト音声合成に応用し，従来の HMM に基づく統計的パラメトリック音声合成の問題点およびノンパラメトリックモデルによるアプローチの有効性を明らかにする。

３．研究の方法

研究目的を達成するために研究課題を具体的な以下の３項目に分け，それぞれについて検討を進める。

(1) ガウス過程回帰モデルに基づく音声合成の基本技術の確立

提案手法では，合成に用いるスペクトルや韻律に関するフレーム単位の音声特徴量を出力(目的変数)とし，入力(説明変数）として言語情報から得られる各フレームの音韻・韻律情報および各フレームの相対位置に関する情報を利用する。本研究ではこれらのフレーム位置に関する情報を「フレームコンテキスト」と呼び，ガウス過程回帰に必要なフレームコンテキストのカーネル関数について検討する。

まず，先行・後続の音韻情報と時刻情報をフレームコンテキストとし，その類似度をカーネルとして利用する。これを音素レベルの音声単位に適用し，音響特徴量のモデル化およびパラメータ生成の基礎的検討を行う。

次に，音素レベルにおける提案手法の有効性を確認した後，これを連続音声へと拡張する。この際のフレームコンテキストとして，先行・後続の音韻情報のみでは不十分であると考えられるため，音韻情報に加えて韻律情報を考慮したコンテキストの拡張を行う。さらに，そのままでは膨大な計算量が必要となることから，スパースカーネルなどを用いた近似による計算量削減の検討を行う。

最終的には，以上の成果に基づいてガウス過程回帰に基づく音声合成（GPR 音声合成）システムのプロトタイプを構築する。

(2) ハイパーパラメータの自動推定とデータ量増加に伴う計算量削減の検討

ガウス過程回帰ではカーネル関数の適切な選択が性能に影響を及ぼし，ガウス過程のハイパーパラメータであるカーネル関数のパラメータをいかに決定するかが重要な問題となる。ここでは，このハイパーパラメータの自動最適化法を検討し，ハイパーパラメータの調整のコストを省くとともに、合成音声の高品質化を目指す。

(3) 多様な音声合成への適用

従来の HMM 音声合成では十分な再現性が得られていないオーディオブックなどに含まれる話し言葉音声や歌唱音声に対し，提案手法を適用しその有効性を明らかにする。このために，オーディオブックなどの評価実験用音声データの収録やラベリングを行い，従来法との比較実験を行う。

なお，本研究は新規性が高い挑戦的課題であることに鑑みて，上記３項目中(1)および(2)に重点を置いて研究を進め，研究の進展状況に応じて項目(3)を可能な範囲で実施することにする。

４．研究成果

(1) ガウス過程回帰に基づく音声合成の基本技術の確立

ガウス過程（GP）は教師あり機械学習に広

く使用されているモデルであり，モデルの複雑さに対する柔軟性と過学習に対する頑健性を兼ね備えたノンパラメトリックベイズモデルとして知られている。本研究で提案するガウス過程回帰（GPR）に基づく音声合成（GPR 音声合成）では，テキストや音声の書き起こしから得られる各フレームの言語特徴量を入力変数，スペクトルなどの各フレームの音響特徴量を出力変数として，ガウス過程の枠組みに基づいて定式化している。提案手法では，フレームレベルのカーネル関数を定義することによって，動的特徴量や木構造のクラスタリングを用いずにフレームの音響特徴量を直接モデル化することを可能としている。

まず初期段階の基礎的検討として，単純なフレームコンテキストカーネルを用いて実験を行った。従来の HMM 音声合成の研究の知見により得られた音素などの音韻情報を時刻情報とともにフレームコンテキストとして用い，そのフレームコンテキストの類似度をカーネルとして利用した。先行・後続の音韻情報と時刻情報のみを考慮した単純フレームコンテキストを用いて音素単位のモデル化と音声合成を行った結果，従来の隠れマルコフモデルに基づく手法に比べて提案法では合成音声のスペクトル歪が減少することを確認した。

音素単位での提案法の有効性が確認できたことを受け，次にフレームコンテキストの拡張と連続音声データへの適用の検討を行った。文章単位の音声合成に用いるコンテキストとして，先行・後続の音素のみでは不十分であると考えられることから，音素境界を考慮したフレームコンテキストの拡張を行った。また，そのままでは膨大な計算量が必要となるため，局所近似や PIC(partially

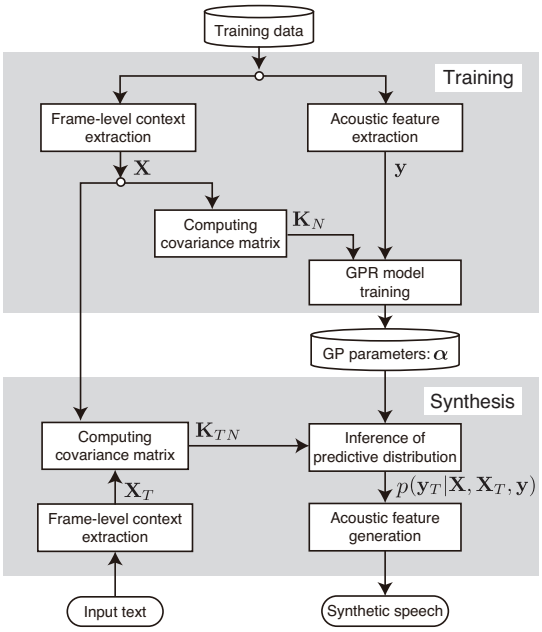


図1　GPR 音声合成の基本構成〔雑誌論文④〕

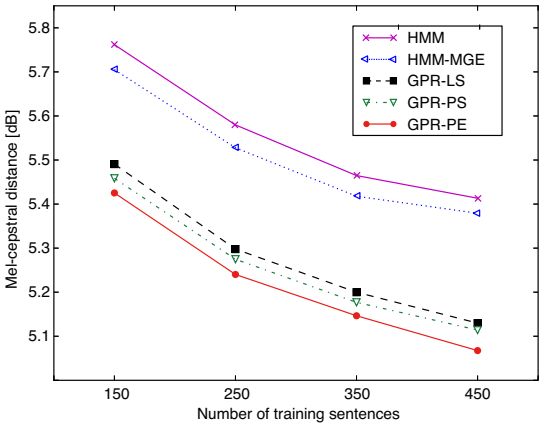


図 2　女性話者の合成音声のスペクトル歪の比較（HMM：HMM 音声合成，HMM-MGE：HMM 音声合成/MGE 学習，GPR-LS/PS/PE：GPR 音声合成，L は局所近似，P は PIC 近似，S は単純フレームコンテキスト，E は拡張フレームコンテキストを表す）〔雑誌論文④〕

independent condition)近似により計算量の削減を行った。

図 1 に示す GPR 音声合成システムを構成し，スペクトル情報を表す音響特徴量のモデル化とパラメータ生成を行った。そして，文章単位の音声合成実験を行い，スペクトル歪に関する客観評価に対して，図 2 に示すように，従来の従来の HMM 音声合成手法を上回る性能が得られることを示した。さらに聴取試験による主観評価でも，HMM 音声合成を上回るスコアが得られた〔雑誌論文④⑤〕〔学会発表⑥〕。

さらに，従来の HMM 音声合成手法において，生成パラメータの過剰平滑化の抑制に有用性が知られている系列内変動（GV）を提案モデル化手法に導入した定式化を行った〔学会発表④〕。その結果，合成音声のスペクトル歪をさらに減少できることを示した。この他にも，フレームコンテキストの選択が合成音声の品質に及ぼす影響について検討を行った〔学会発表①〕。

(2) ハイパーパラメータの自動推定とデータ量増加に伴う計算量削減の検討

提案する GPR 音声合成では，適切なカーネル関数の選択およびカーネル関数のパラメータの決定が重要な問題の一つとなる。実際に，GPR による予測精度はある程度カーネル関数に依存しており，データに適したカーネル関数を使用することが望ましい。さらに，カーネル関数のパラメータを勾配法などで自動的に最適化することが可能であるとはいえ，提案する GPR 音声合成の枠組みでは，勾配法の各ステップにおいて大量の計算量が必要になってしまう．

これに対し，ハイパーパラメータであるカーネル関数のパラメータとノイズパラメータを効率的に最適化する手法を提案した〔雑

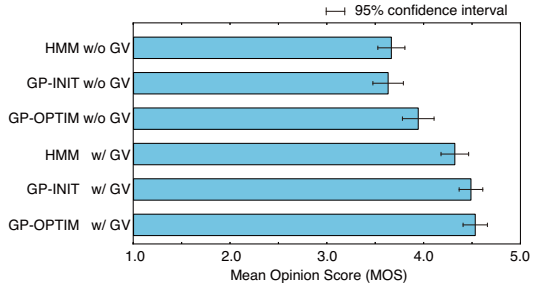


図 3 合成音声の自然性の観点から評価したハイパーパラメータ最適化の効果（HMM：HMM 音声合成，GP-INIT：GPR 音声合成・最適化前，GP-OPTIM：GPR 音声合成・最適化後，w/o GV：GV 考慮なし，w/ GV：GV 考慮）〔雑誌論文③〕

誌論文③〕〔学会発表⑤〕。図 3 に GPR 音声合成システムにおけるハイパーパラメータの最適化前後の合成音声の主観評価スコアを示す。比較のため，HMM 音声合成，GV の考慮の有無の場合についてもスコアを示してある。図に示す通り，ハイパーパラメータの自動最適化により，有意に自然性が向上する結果が得られた。

この他に，提案手法の実用化の観点から，音声合成パラメータ生成時における計算量と予測性能の間のトレードオフやフットプリントに関しても詳細な検討を行った〔雑誌論文②〕。

(3) 多様な音声合成への適用

多様な音声合成に提案手法である GPR 音声合成を適用するためには，音声のスペクトル情報に加えて韻律情報のモデル化・パラメータ生成の枠組みの開発が必要である。これに対し，ガウス過程分類（GPC）を利用した有声／無声区間推定，GPR に基づく基本周波数パタンのモデル化とパラメータ生成手法を提案した〔学会発表③〕。さらに，これらをスペクトル情報のモデル化・パラメータ生成と組み合せ，提案 GPR 音声合成手法のプロトタイプシステムを構築した〔雑誌論文①〕〔学会発表②〕。

図 4 に読上げ調音声を対象とし，提案 GPR 音声合成システムにより韻律を含めて合成した音声と，従来法である HMM 音声合成により得られた音声を自然性の観点から対比較評価を行った結果を示す。この結果より，提案手法は HMM 音声合成手法の限界を超える性能を持つ可能性があることが確認できた。

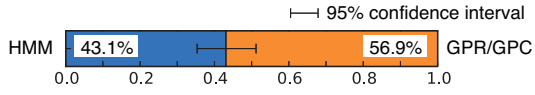


図 4 合成音声の自然性の主観評価結果（HMM：HMM 音声合成，GPR/GPC：GPR 音声合成）〔雑誌論文①〕

この他に，読上げ調音声と比べて合成音声の再現がより難しいオーディオブック向け音声と歌唱音声の収録を行い，提案手法の性能評価を行うための基盤整備を行った。なお，これらの音声を対象とした提案手法の詳細の評価は今後の課題とする。

本研究で得られた成果を基に，今後は GPR 音声合成のプロトタイプシステムを実用的なシステムへと発展させると共に，多様な話者性や話し言葉を含む多様なスタイルによる音声合成，多言語音声合成へと，研究を展開して行く予定である。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 11 件）


① Tomoki Koriyama, Takao Kobayashi, Prosody generation using frame-based Gaussian process regression and classification for statistical parametric speech synthesis, 査読有, Proc. 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2015, pp.4929-4933, 2015.

② Tomoki Koriyama, Takashi Nose, Takao Kobayashi, Parametric speech synthesis using local and global sparse Gaussian processes, 査読有, Proc. 2014 IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2014, pp.1-6, DOI: 10.1109/MLSP.2014.6958921, 2014.

③ Tomoki Koriyama, Takashi Nose, Takao Kobayashi, Parametric speech synthesis based on Gaussian process regression using global variance and hyperparameter optimization, 査読有, Proc. 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2014, pp.3862-3866, DOI: 10.1109/ICASSP.2014.6854319, 2014.

④ Tomoki Koriyama, Takashi Nose, Takao Kobayashi, Statistical parametric speech synthesis based on Gaussian process regression, 査読有, IEEE Journal of Selected Topics in Signal Processing, Vol.8, pp.173-183, DOI: 10.1109/JSTSP.2013.2283461, 2014.

⑤ Tomoki Koriyama, Takashi Nose, Takao Kobayashi, Statistical nonparametric speech synthesis using sparse Gaussian processes, 査読有, Proc. 14th Annual Conference of the International Speech Communication Association, INTERSPEECH 2013, pp.1072-1076, http://www.isca-speech.org/archive/archive_papers/interspeech_2013/i13_1072.pdf, 2013.

〔学会発表〕(計 10 件)

① 岡元伶洋, ガウス過程回帰に基づく音声合成のためのコンテキストの検討, 日本音響学会 2015 年春季研究発表会, 2015 年 3 月 17 日, 中央大学(東京都文京区).
② 郡山知樹, ガウス過程回帰に基づく音声合成システムの検討, 日本音響学会 2014 年秋季研究発表会, 2015 年 3 月 17 日, 中央大学(東京都文京区).
③ 郡山知樹, ガウス過程回帰に基づく F0 パタン生成の検討, 日本音響学会 2014 年秋季研究発表会, 2014 年 9 月 4 日, 北海学園大学(北海道札幌市).
④ 郡山知樹, 系列内変動を考慮したガウス過程回帰に基づく音声パラメータ生成, 日本音響学会 2014 年春季研究発表会, 2014 年 3 月 12 日, 日本大学(東京都千代田区).
⑤ 郡山知樹, ガウス過程回帰に基づく音声合成におけるハイパーパラメータ最適化の検討, 電子情報通信学会・日本音響学会音声研究会, 2014 年 1 月 23 日, 名城大学, (愛知県名古屋市).
⑥ 郡山知樹, スパース近似と畳み込みカーネルを用いたガウス過程回帰に基づく音声合成, 日本音響学会 2013 年秋季研究発表会, 2013 年 9 月 26 日, 豊橋技術科学大学(愛知県豊橋市).

〔その他〕
ホームページ等
http://www.kbys.ip.titech.ac.jp/

６．研究組織

(1)研究代表者
小林　隆夫 (KOBAYASHI, Takao)
東京工業大学・大学院総合理工学研究科・教授
研究者番号：70153616

(2)研究分担者
能勢　隆 (NOSE, Takashi)
東北大学・大学院工学研究科・講師
研究者番号：90550591

(3)連携研究者

(4)研究協力者
郡山　知樹 (KORIYAMA, Tomoki)
東京工業大学・大学院総合理工学研究科・助教
研究者番号：50749124