

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Prosody Modeling Based on Gaussian Process Regression for Thai Speech Synthesis
著者(和文)	MoungsriDecha
Author(English)	Decha Moungsri
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10923号, 授与年月日:2018年6月30日, 学位の種別:課程博士, 審査員:小林 隆夫,奥村 学,山口 雅浩,杉野 暢彦,篠崎 隆宏
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10923号, Conferred date:2018/6/30, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

報告番号	甲第		号	学位申請者氏名		Decha Mungsri	
		氏名		職名		氏名	職名
論文審査 審査員	主査	小林 隆夫		教授	審査員	篠崎 隆宏	准教授
	審査員	奥村 学		教授			
		山口 雅浩		教授			
		杉野 暢彦		准教授			

論文審査の要旨 (2000 字程度)

本論文は「Prosody Modeling Based on Gaussian Process Regression for Thai Speech Synthesis (タイ語音声合成のためのガウス過程回帰に基づく韻律モデル化)」と題し、英文7章からなっている。

第1章「Introduction (序論)」では、まず研究背景として、スマートスピーカやスマートフォン上で利用されている音声インタフェースに欠かせない要素技術としてテキスト音声合成があり、従来は隠れマルコフモデル (HMM) に基づいた音声合成手法が優位であったが、より自然性の高い合成音声生成への要求から、最近では深層学習や機械学習を導入した手法が主流になっていることを述べ、次に声調言語であるタイ語の音声合成では、声調や音韻継続長を含む韻律特徴が自然性を左右する要因であり、より自然なタイ語音声合成の実現をめざして、機械学習手法の一つであるガウス過程回帰 (Gaussian process regression, GPR) を導入した新たな韻律モデル化手法を確立することが本研究の目的であると述べている。

第2章「Gaussian Process Regression-Based Thai Speech Synthesis (ガウス過程回帰に基づくタイ語音声合成)」では、本研究で対象とするガウス過程回帰に基づく音声合成 (GPR 音声合成) について、タイ語の韻律特徴を概説した後、システム構成、使用するコンテキストやカーネル関数を定義した上で、GPR 音声合成システムの基本性能の評価を行い、学習音声データ量が1ないし2時間程度の場合、従来の HMM 音声合成や深層ニューラルネット (DNN) に基づく DNN 音声合成に比べて客観評価、主観評価ともにより高い性能を持つことを明らかにしている。

第3章「Two-Stage GPR-Based Duration Prediction (2段 GPR 継続長予測)」では、より正確な音素継続長予測をめざし、音節単位の継続長が自然性に影響を及ぼすことを考慮して、まず音節継続長を予測するモデルを作り、次にこれから得られた予測値を新たなコンテキストとして加えて音素継続長を予測するという、2段構成のガウス過程回帰に基づく継続長予測手法を提案している。さらに、音素継続長予測を2段構成にするアプローチを従来の HMM 音声合成や DNN 音声合成の枠組みにも適用し、それぞれの音声合成の枠組みを用いて詳細な予測性能の比較評価を行っている。その結果、提案手法が客観評価、主観評価とも従来手法に比べて有意に評価結果が上回ることを明らかにしている。

第4章「Duration Prediction Using Multiple Gaussian Process Experts for GPR-Based Speech Synthesis (GPR 音声合成のための複数ガウス過程エキスパートを用いた継続長予測)」では、前章とは異なる複数レベルの継続長を考慮した音素継続長予測手法を検討している。前章で述べた手法では、音節継続長予測値をコンテキストの形で間接的に用いて音素継続長予測をしていたのに対し、ここでは音節継続長と音素継続長の二つの予測モデルの積により表されたモデルから最適な音素継続長を予測している。約1時間の学習音声データを用いた評価結果より、音素継続長単独予測モデルと比べてより性能が高いこと、前章で述べた手法との比較では、主観評価ではほぼ同等であることが示されている。

第5章「Enhanced F0 Generation for GPR-Based Speech Synthesis Combining Syllable-Based Prosodic Features (音節単位の韻律特徴を組み合わせた GPR 音声合成のための改良 F0 生成)」では、基本周波数 (F0) パタン生成時に音節単位の韻律特徴を考慮した手法を提案している。ここでは、音節単位の F0 パタンを表す離散コサイン変換 (DCT) 係数を出力するガウス過程モデルを作り、これを通常の GPR 音声合成の F0 パタン生成モデルと結合して F0 パタン生成を行っている。評価実験の結果、少量の学習データしか使用できない場合に提案手法の有効性が示されている。

第6章「Unsupervised Stress Information Labeling Using Gaussian Process Latent Variable Model for Statistical Speech Synthesis (統計的音声合成のためのガウス過程潜在変数モデルを用いた教師なしストレス情報ラベリング)」では、タイ語の韻律がストレスの有無により大きく変化すること、ストレスの有無は入力テキスト情報のみからでは得られないことから、韻律モデル学習において学習データに対し適切にストレス情報ラベルを付与することが重要であることを述べ、その解決策として、ガウス過程潜在変数モデルを用いた教師なし学習手法を提案し、評価実験を通してその有効性を示している。

第7章「Conclusions (結論)」では、本論文の成果をまとめ、今後の課題を述べている。

以上を要するに、本論文では、タイ語音声合成において合成音声の自然性を左右する要因である韻律生成に関し、ガウス過程回帰に基づく新たなモデル化手法を提案し、評価実験を通してその有効性を明らかにしたものであり、工学上ならびに工業上寄与するところが大きい。よって本論文は、博士 (工学) の学位論文として価値あるものと認められる。