

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Statistical Speech Synthesis Using Extended Context and Gaussian Process Regression
著者(和文)	郡山知樹
Author(English)	Tomoki Koriyama
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9323号, 授与年月日:2013年9月25日, 学位の種別:課程博士, 審査員:小林 隆夫,羽鳥 好律,伊東 利哉,小池 康晴,篠崎 隆宏
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9323号, Conferred date:2013/9/25, 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		郡山 知樹	
		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	羽鳥 好律	教授				
		伊東 利哉	教授				
		小池 康晴	教授				

論文審査の要旨 (2000 字程度)

本論文は「Statistical Speech Synthesis Using Extended Context and Gaussian Process Regression (拡張コンテキストおよびガウス過程回帰を用いた統計的音声合成)」と題し、英文6章からなっている。

第1章「Introduction (序論)」では、本研究の背景として、入力テキストからその読上げ音声コンピュータにより合成するテキスト音声合成の研究ならびに当該研究分野で近年主流となりつつある統計的パラメトリック音声合成について述べ、統計的パラメトリック音声合成手法の一つである隠れマルコフモデル (hidden Markov model, HMM) に基づく音声合成 (HMM 音声合成) では、日常生活における普通の会話等の自発性の高い音声を対象とした合成音声の韻律の自然性やスペクトルの再現性が不十分であり、これを解決するための新たな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「Statistical Parametric Speech Synthesis (統計的パラメトリック音声合成)」では、本研究で提案する新たなモデル化手法のベースラインとなる、HMM 音声合成の基本を概説している。

第3章「Spontaneous Speech Synthesis Using Extended Context (拡張コンテキストを用いた自然発話音声合成)」では、自然発話会話音声合成に向けて、多様な韻律を生成するためのコンテキストに関する検討を行っている。HMM 音声合成では、スペクトルや韻律の変動を詳細にモデル化するために、スペクトルや韻律に影響を及ぼす変動要因 (コンテキスト) を考慮したコンテキスト依存モデルを用いている。従来のコンテキストは、主に入力テキストから得られる言語情報に基づいており、これをそのまま自然発話会話音声合成に適用しても多様な韻律を実現できなかった。これに対し、日本語話し言葉コーパス (CSJ) に付与されている様々な情報を拡張コンテキストとして用いる手法を提案し、さらにコンテキストの増加によるモデル過学習を避けるための手法を提案している。提案手法を CSJ に含まれる対話音声に適用し、合成音声の客観及び主観評価実験を通して、音素引き延ばしと基本周波数 (F0) 変動の種類を表すトーンラベルが拡張コンテキストとして有効であることを示し、さらに拡張コンテキストを自動的に求める手法の検討を行い、正解コンテキストを用いた場合と同程度の自然性が得られることを明らかにしている。

第4章「Prosodic-Event-Based HMM (韻律イベントHMM)」では、自然発話会話音声の F0 をより効率よくモデル化するための新たなモデル単位を検討している。従来の HMM 音声合成では、F0 のモデル化を音素単位で行っているが、F0 特徴の変化は必ずしも個々の音素境界と同期しているわけではないため、音素単位の F0 モデル化は効率的とは言えない。これに対し、アクセント位置における F0 の下降やアクセント句末に生じる F0 変動等の韻律イベントを境界とし、この境界間のセグメントを単位とする新たな F0 モデル化手法を提案している。CSJ に含まれる男女各1名の対話音声を用いて客観及び主観評価実験を行い、音素単位の F0 モデル化に比べ、提案手法はモデルパラメータ数が半分以下にもかかわらず、ほぼ同等の客観・主観性能が得られること、F0 変動の多様性がより自然発話音声に近づくことを示し、効率的なモデル化が実現できることを明らかにしている。

第5章「Speech Synthesis Based on Gaussian Process Regression (ガウス過程回帰に基づく音声合成)」では、HMM に基づくモデル化手法の限界を超える新たな統計的モデル化手法について検討している。一般に HMM では、複数の分析フレームをまとめて一つの状態に割当て、同一状態内ではモデルパラメータが変化しないことを仮定している。これに対し、ガウス過程回帰を HMM の代わりに用いる手法を提案し、学習データ量に応じたより精密なフレーム単位でのモデル化を可能にしている。ここでは、提案手法をスペクトルパラメータのモデル化に適用し、カーネルの設計、音声合成のためのパラメータ生成、計算量の削減等の各手法について述べ、提案手法の評価の第一歩として、読上げ音声を用いた客観及び主観評価実験の結果、提案手法は HMM に比べて同等かそれ以上の性能が得られることを明らかにしている。

第6章「Conclusions and Future Work (結論と今後の課題)」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、本論文では自然発話音声の合成に向けて、韻律の多様性やスペクトルの自然性が高い音声を合成するためのいくつかの新たな枠組みを提案し、従来手法との比較評価実験を通してその有効性と可能性を明らかにしたものであり、工学上ならびに工業上貢献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。