

論文 / 著書情報  
Article / Book Information

題目(和文)	
Title(English)	Statistical Parametric Speech Synthesis Using Local Variance and Quantized F0 Context
著者(和文)	CHUNWIJITRAVATAYA
Author(English)	Vataya Chunwijitra
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9322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,小池 康晴,杉野 暢彦,篠崎 隆宏
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9322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

(博士課程)

## 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		Chunwijitra, Vataya	
論文審査 審査員		氏名	職名		氏名	職名
	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			
		小池 康晴	教授			
杉野 暢彦		准教授				

### 論文審査の要旨 (2000 字程度)

本論文は「Statistical Parametric Speech Synthesis Using Local Variance and Quantized F0 Context (局所変動と量子化 F0 コンテキストを用いた統計的パラメトリック音声合成)」と題し、英文 5 章からなっている。

第 1 章「Introduction (序論)」では、人間とコンピュータ間のインタフェースとしてのテキスト音声合成に関する近年の研究・実用化動向について述べ、音声のパラメトリック表現と隠れマルコフモデル (hidden Markov model, HMM) に基づいた統計的パラメトリック音声合成手法が、合成音声の多様性、システム実現における容易さ、コスト等の観点から注目されているものの、声調 (トーン) 言語、特にタイ語のように音声資源が十分に整備されていない言語の音声合成ではトーン其自然性が十分でない場合があること、またすべての言語に共通して HMM に基づく統計的モデル化では、パラメータの過剰平滑化に起因する合成音品質の劣化が起こるという問題点を指摘した上で、これらの問題点を解決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、本研究で対象とする統計的パラメトリック音声合成手法の隠れマルコフモデルに基づく音声合成 (HMM 音声合成) について、音声合成単位のモデル化、システム構成、コンテキスト依存モデル、平均声モデル等について概説した上で、第 1 章で指摘した声調言語の音声合成におけるトーン其自然性と、パラメータの過剰平滑化に起因するスペクトルの再現性の問題点に関して、これまでに提案されているアプローチを解説している。

第 3 章「Tone-Modeling Using a Quantized F0 Context in Average-Voice-Based Speech Synthesis (平均声に基づく音声合成における量子化 F0 コンテキストを用いたトーンのモデル化)」では、多数話者の音声データから学習された平均声モデルと話者適応を用いて任意の目標話者の音声を合成するシステムにおいて、声調言語を対象としたトーンの新たなモデル化手法について述べている。平均声モデルを学習する際に、プロのナレータやアナウンサーでない一般話者の音声データを用いた場合、トーンを特徴付ける基本周波数 (F0) の時間変化パターンと、発話テキストの言語情報から定まるトーンの型が必ずしも一致しないことがしばしば起こり、この不一致が音声合成時におけるトーンの不自然さの要因となっている。そこで、トーンのモデル化においてコンテキスト依存モデルを作成する際、従来のトーンの型をコンテキストとして用いる代わりに、正規化された対数 F0 の値を量子化して得られた F0 シンボルをコンテキストとする手法を提案している。男女各 24 名、合計 48 名の一般話者が各 38 文章を発話したタイ語音声データから平均声モデルを作成し、平均声方式により男女各 2 名を目標話者として合成音声のトーンの正確さと自然性を評価した結果、提案手法は客観評価、主観評価ともに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「Parameter Generation Using Local Variance for HMM-Based Speech Synthesis (HMM 音声合成のための局所変動を用いたパラメータ生成)」では、HMM を用いたスペクトルパラメータのモデル化において、局所変動 (local variance, LV) パラメータを新たに特徴量に加えたモデル化と、得られたモデルからの LV を考慮したスペクトルパラメータ生成手法を提案している。従来、HMM によるモデル化に伴う過剰平滑化の問題を緩和する手法として、文章単位でのパラメータの大域変動 (global variance, GV) 特徴量を考慮したモデル化とパラメータ生成手法が良く用いられているが、生成されたスペクトルの歪の増加やパラメータの過剰強調の問題があった。これに対し提案手法は、スペクトル歪の増加やパラメータの過剰強調を押さえつつ、GV 特徴量を考慮する場合と同等かそれ以上の聴感上の品質向上に成功している。日本語を対象として、プロのナレータによる男女各 1 名の音声データを用いたモデル化と合成音声の評価実験の結果から、提案手法の有効性を示している。

第 5 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、統計的パラメトリック音声合成手法の HMM 音声合成において、声調言語のトーン其自然性を向上させる手法と、言語によらず合成音声のスペクトルの再現性を向上させる手法を提案し、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。