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)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	物理情報システム　専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of　（ 工学 ）
学籍番号： Student ID Number		指導教員（主）： Academic Advisor(main)	小林　隆夫
学生氏名： Student's Name	郡山　知樹	指導教員（副）： Academic Advisor(sub)	

要旨（和文 2000 字程度）
Thesis Summary（approx.2000 Japanese Characters ）

本論文は統計的音声合成の枠組みにおいて，多様な韻律および高い自然性を持つ音声を合成する手法を提案する．近年，音声合成の応用場面が広がるにつれて，音声合成システムにも様々な機能が求められるようになっている．これまでの研究によって，合成音声の多様性および自然性は向上しつつあるが，人間のような自然な音声を生成するには，未だ至っていない．本論文では，多様な韻律表現を持つ話し言葉対話音声に注目し，自然な話し言葉音声を合成する枠組みを実現することによって，合成音声の多様性を向上させることを目標としている．

まず，統計的音声合成として広く利用されている HMM 音声合成の枠組みにおいて，韻律の多様性を実現するために，拡張コンテキストの導入を行う．コンテキストとは音韻や韻律の特徴を表す変動要因であるが，本研究では日本語話し言葉コーパスから得られるコンテキストを従来のコンテキストに加えることで，有効性の評価を行う．客観評価実験の結果，音素引き延ばしとトーンラベルに関するコンテキストが，基本周波数(F0)および継続長のモデル化に有効であることがわかった．また，主観評価実験においても拡張コンテキストは従来のコンテキストに比べ，より自然な話し言葉音声を生成することが確認された．

次に，話し言葉対話音声の韻律の自然性を向上させるため，韻律イベント HMM を提案する．韻律イベント単位 HMM のモデル化単位は，従来の HMM 音声合成における音素と異なり，アクセントによるピッチの下降や句末境界音調におけるピッチ上昇などの韻律イベントを表すトーンラベルに基づいている．韻律イベントはスペクトル情報よりも F0 情報に基づくイベントであるため，提案法である韻律イベント HMM は F0 に関するモデルパラメータ数を抑えられることを期待できる．実験結果より，提案法は従来の音素単位の HMM に比べ，よりコンパクトなモデルを生成し，また，変動の大きい F0 を生成することが示された．

合成音声の韻律の多様性および自然性が拡張コンテキストおよび韻律イベント HMM によって改善されるとはいえ，スペクトル特徴量の自然性に関しては依然として不十分である．そこで，本論文ではスペクトル特徴量の自然性向上のためガウス過程回帰に基づく統計的音声合成の枠組みを提案する．音声パラメータのトラジェクトリのモデル化を実行可能な時間で計算するため，local GPs や PIC といったスパース近似による近似を用いる．さらに，滑らかな音声パラメータのトラジェクトリを生成するために，隣接する音素の情報を含むコンテキストを導入し，畳み込みカーネルを用いて各フレームの入力変数のカーネル関数を定義する．客観評価および主観評価の結果から，提案法は比較的自然なスペクトル特徴量を生成可能なことが示された．

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 2 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 2 copies of 800 Words (English).

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	物理情報システム　専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of　（　工学　）
学籍番号： Student ID Number		指導教員（主）： Academic Advisor(main)	小林　隆夫
学生氏名： Student's Name	郡山　知樹	指導教員（副）： Academic Advisor(sub)	

要旨（英文 300 語程度）
Thesis Summary（approx.300 English Words ）

This thesis describes novel approaches for synthesizing speech with prosodic variability and naturalness. There are a variety of applications of speech synthesis and there have been increasing demands for such applications. Although the variability and naturalness of synthetic speech have been improved, the ability of generating natural sounding speech is still insufficient. This thesis focuses on spontaneous conversational speech that has much prosodic variability. The purpose of this study is to improve the variability using spontaneous speech data by realizing the framework that can synthesizes natural-sounding spontaneous conversational speech.

First, extended context is introduced to synthesize natural-sounding spontaneous conversational speech with prosodic variability in the hidden-Markov-model-based speech synthesis framework. Several context sets that can be obtained from the Corpus of Spontaneous Japanese are introduced and the effectiveness of the context sets is evaluated. The results of objective evaluation show that the phone prolongation and tone labels are effective for improving generated F0 and duration. It has been confirmed that the synthetic speech using extended context offers more natural-sounding speech than conventional contexts from the subjective evaluation.

Next, prosodic-event-based HMM (prosodic-unit HMM) is proposed to improve the naturalness of prosody of spontaneous conversational speech. The modeling unit proposed prosodic-event-based HMM is the segment between two tone labels that represents prosodic events such as pitch falling by accent or pitch rising of boundary pitch movement (BPM). The proposed HMM is expected to reduce the model parameters of F0 because there are less prosodic events derived from F0 features than phones that strongly depends on spectral features. The results show that the proposed technique gives a more compact model and more variation in generated F0 than phone-unit HMM.

The prosodic variability and naturalness of synthetic speech is improved by extended context and prosodic-event-based HMM. However the naturalness of spectral features is still insufficient. Then, a speech synthesis framework based on Gaussian process regression is proposed to improve the naturalness of spectral features. Block-based sparse GP approximations such as local GPs and PIC are used for trajectory modeling of utterances with feasible computation. Moreover, for the generation of smooth parameter trajectory, frame context including nearby phone information and its kernel is defined. From the objective and subjective evaluation, the proposed method using the PIC approximation and the extended context achieved better performance than the HMM-based methods.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 2 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 2 copies of 800 Words (English).