

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 9 日現在

機関番号：11301

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700195

研究課題名(和文)次世代ヒューマノイド音声合成に向けたコーパスデザインに基づく音声多様化技術の研究

研究課題名(英文)A study on speech diversification techniques based on corpus design for advanced humanoid speech synthesis

研究代表者

能勢 隆(Nose, Takashi)

東北大学・工学(系)研究科(研究院)・講師

研究者番号：90550591

交付決定額(研究期間全体)：(直接経費) 3,300,000円、(間接経費) 990,000円

研究成果の概要(和文)：本研究課題は人間に近いより自然で多様な音声合成システムの実現を目指すものであり、研究期間全体を通じて以下の項目について研究成果が得られた。

(1)アクセント、スタイル、文末表現などを総合的に考慮した音声コーパス構築法を提案し、その有効性を示した。(2)ユーザによる主観的な感情の度合を定量化してモデル学習に組み込む新たな枠組を提案した。(3)強調表現を含む音声を自動的に生成するため、声の高さを表す基本周波数(F0)に着目し、F0生成を利用した強調表現の自動ラベリングを実現した。(4)言語の多様化手法として、目標話者の母国語音声のみからその話者の外国語音声合成する手法を提案した。

研究成果の概要(英文)：Our goal in this research is to realize more human-like, natural text-to-speech system with various emotional expressions and speaking styles, and the achievements of our studies are as follows:

(1)We proposed a novel corpus-design technique in which accent, style, and sentence-final expression are taken into account. (2)We incorporated user's subjective emotional intensities into acoustic model training to improve the performance of expressive speech synthesis. (3)We proposed an automatic labeling technique of emphasis expression using a parameter generation technique of fundamental frequency to realize emphatic speech synthesis. (4)We proposed cross-lingual speech synthesis using only a target speaker's native language speech samples to synthesis multi-lingual speech at a low cost.

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声合成 隠れマルコフモデル 統計的音声合成 感情音声合成 ヒューマノイドロボット 音声コーパス

### 1. 研究開始当初の背景

政府の公約である長期的戦略指針「イノベーション25」にも掲げられるように、人間と自然なコミュニケーションをとりながらサポートを行うロボットの普及に向けての取り組みが進められている。この実現のためには人間に近い対話機能が必要となり、音声合成はそのための重要な基盤技術の一つである。近年、計算機の性能向上、合成技術の進歩によって、アナウンサーなどの読み上げ調の音声については品質が大きく向上している。一方で、ユーザにとってより自然で違和感のない音声対話インターフェースを実現するためには、人間と同様に、多種多様な感情、発話様式、発話意図などのパラ言語を適切に表現・変化させることができる高品質な音声合成技術が必須となるが、未だ実用化には至っていない。

### 2. 研究の目的

本研究では、研究代表者らがこれまでに提案してきた、感情や発話様式を表現・制御可能な音声合成手法をさらに発展させ、あたかも人間が話しているかのような、より多くの感情・発話様式・発話意図が自然かつ多様に变化する次世代の高品質音声合成システムの実現を目指す。具体的には

- (1) 実環境での利用において必要とされる様々な感情・発話様式・発話意図などを含んだ音声合成向けパラ言語音声データベースの分析・構築
- (2) 上記の音声コーパスを用いた、多様なパラ言語表現が入り交じった高品質で人間らしい音声合成技術の確立

の2項目について研究を進める。

### 3. 研究の方法

研究期間は3年間であり、研究目的の達成のため具体的な研究を以下の4項目に分け、それぞれについて実施した。

- (1) 模擬/自然発話による対話音声の収録とパラ言語情報の分析
- (1) 対話音声の分析に基づくパラ言語音声コーパスのデザインと構築
- (2) コンテキストの多様化による多様なパラ言語情報を含んだ音声合成
- (2) 重回帰 HMM に基づく合成音声の多様化・高品質化

初年度(平成23年度)は、項目(1)および(2)の対話音声の分析および音声コーパスの構築に重点をおいて研究を進めた。そして次年度(平成24年度)以降は構築したコーパスを利用し、項目(1)および(2)の合成音声の多様化・高品質化技術の開発に重点をおいて取り組んだ。

### 4. 研究成果

本研究課題は、人間に近いより自然で多様な音声合成システムの実現を目指すものであり、研究期間全体を通じて以下の項目について研究成果が得られた。

- (1) 対話音声合成のための効率的なコーパスデザイン法の確立

アクセント、スタイル、文末表現などを総合的に考慮した音声コーパス構築法を提案し、その有効性を示した。

- (2) ユーザの主観に基づいた感情音声合成  
ユーザによる主観的な感情の度合いを定量化してモデル学習に組み込む新たな枠組を提案し、その有効性を示した。

- (3) 話し言葉音声合成のための韻律モデルの改善

韻律の変動要因として ToBI ラベルに基づく詳細な情報を導入した韻律単位 HMM を提案し、その有効性を示した。

- (4) 教師なし学習に基づく強調音声の合成  
強調表現を含む音声を自動的に生成するため、声の高さを表す基本周波数(F0)に着目し、F0生成を利用した強調表現の自動ラベリングを実現した。

- (5) 多様なスタイル音声生成のためのスタイル変換法の提案

複数の話者の通常音声と感情音声の間で変換行列を推定しこれを目標話者の通常音声のモデルに適用することで目標話者の読み上げ音声のみから多様なスタイルを伴う音声を生成できることを示した。

- (6) 歌声合成におけるスタイル制御法の提案  
スタイルとその度合いを直観的に変化させることができる重回帰隠れセミマルコフモデルに基づく歌唱スタイル制御法を提案した。

- (7) 共有決定木に基づくクロスリンガル音声合成

言語の多様化手法として、目標話者の母国語音声のみからその話者の外国語音声を合成する手法を提案した。

### 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20件)

- (1) Yu Maeno, Takashi Nose, Takao Kobayashi, Tomoki Koriyama, Yusuke Ijima, Hideharu Nakajima, Hideyuki Mizuno, Osamu Yoshioka, "Prosodic variation enhancement using unsupervised context labeling for HMM-based expressive speech synthesis,"

Speech Communication, 査読有, Vol. 57, pp. 144-154, 2014 年  
10.1016/j.specom.2013.09.014

(2) Tomohiro Nagata, Hiroki Mori, Takashi Nose, "Robust estimation of multiple-regression HMM parameters for dimension-based expressive dialogue speech synthesis," Proceedings of 14th Annual Conference of the International Speech Communication Association (ISCA), 査読有, pp.1549-1553, 2013 年

(3) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "Statistical nonparametric speech synthesis using sparse Gaussian processes," Proceedings of 14th Annual Conference of the International Speech Communication Association (ISCA), 査読有, pp.1072-1076, 2013 年

(4) Takashi Nose, Misa Kanemoto, Tomoki Koriyama, Takao Kobayashi, "A style control technique for singing voice synthesis based on multiple-regression HSMM," Proceedings of 14th Annual Conference of the International Speech Communication Association (ISCA), 査読有, pp.378-382, 2013 年

(5) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "Frame-level acoustic modeling based on Gaussian process regression for statistical nonparametric speech synthesis," Proceedings of 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing, 査読有, pp. 8007-8011, 2013 年

(6) Hiroki Kanagawa, Takashi Nose, Takao Kobayashi, "Speaker-independent style conversion for HMM-based expressive speech synthesis," Proceedings of 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing, 査読有, pp. 7864-7868, 2013 年

(7) Yu Maeno, Takashi Nose, Takao Kobayashi, Tomoki Koriyama, Yusuke Ijima, Hideharu Nakajima, Hideyuki Mizuno, Osamu Yoshioka, "HMM-based expressive speech synthesis based on phrase-level F0 context labeling," Proceedings of 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing, 査読有, pp. 7859-7863, 2013 年

(8) Takashi Nose, Takao Kobayashi, "An intuitive style control technique in HMM-based expressive speech synthesis

using subjective style intensity and multiple-regression global variance model," Speech Communication, 査読有, Vol. 55, No. 2, pp. 347-357, 2013 年  
10.1016/j.specom.2012.09.003

(9) Vataya Chunwijitra, Takashi Nose, Takao Kobayashi, "A speech parameter generation algorithm using local variance for HMM-based speech synthesis," Proceedings of 13th Annual Conference of the International Speech Communication Association (ISCA), 査読有, pp.1151-1154, 2012 年

(10) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "Discontinuous observation HMM for prosodic-event-based F0 generation," Proceedings of 13th Annual Conference of the International Speech Communication Association (ISCA), 査読有, pp.462-465, 2012 年

(11) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "An F0 modeling technique based on prosodic events for spontaneous speech synthesis," Proceedings of 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), 査読有, pp. 4589-4593, 2012 年

(12) 郡山知樹, 能勢 隆, 小林隆夫, "HMMに基づく対話音声合成における多様な韻律生成のためのコンテキストの拡張," 電子情報通信学会論文誌, 査読有, Vol. J95-D, No. 3, pp. 597-607 2012 年

(13) Takashi Nose, Takao Kobayashi, "Very low bit-rate F0 coding for phonetic vocoders using MSD-HMM with quantized F0 symbols," Speech Communication, 査読有, Vol. 54, No. 3, pp. 384-392, 2012 年  
10.1016/j.specom.2011.10.002

(14) Vataya Chunwijitra, Takashi Nose, Takao Kobayashi, "A tone-modeling technique using a quantized F0 context to improve tone correctness in average-voice-based speech synthesis," Speech Communication, 査読有, Vol. 54, No. 2, pp. 245-255, 2012 年  
10.1016/j.specom.2011.08.006

(15) Takashi Nose, Takao Kobayashi, "Recent development of HMM-based expressive speech synthesis and its applications," Proceedings of 2011 Asia-Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference, 査読有, 2011 年

[http://www.apsipa.org/proceedings\\_2011/pdf/APSIPA189.pdf](http://www.apsipa.org/proceedings_2011/pdf/APSIPA189.pdf)

(16) Takashi Nose, Takao Kobayashi, "Speaker-independent HMM-based voice conversion using adaptive quantization of the fundamental frequency," *Speech Communication*, 査読有, Vol. 53, No. 7, pp. 973-985, 2011年

10.1016/j.specom.2011.05.001

(17) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "On the use of extended context for HMM-based spontaneous conversational speech synthesis," *Proceedings of 12th Annual Conference of the International Speech Communication Association (ISCA) (INTERSPEECH 2011)*, 査読有, pp. 2657-2660, 2011年

(18) Tatsuhiko Saito, Takashi Nose, Takao Kobayashi, Yohei Okato, Akio Horii, "Performance prediction of speech recognition using average-voice-based speech synthesis," *Proceedings of 12th Annual Conference of the International Speech Communication Association (ISCA) (INTERSPEECH 2011)*, 査読有, pp. 1953-1956, 2011年

(19) Yu Maeno, Takashi Nose, Takao Kobayashi, Yusuke Ijima, Hideharu Nakajima, Hideyuki Mizuno, Osamu Yoshioka, "HMM-based emphatic speech synthesis using unsupervised context labeling," *Proceedings of 12th Annual Conference of the International Speech Communication Association (ISCA) (INTERSPEECH 2011)*, 査読有, pp. 1849-1852, 2011年

(20) Takashi Nose, Takao Kobayashi, "A perceptual expressivity modeling technique for speech synthesis based on multiple-regression HSMM," *Proceedings of 12th Annual Conference of the International Speech Communication Association (ISCA) (INTERSPEECH 2011)*, 査読有, pp. 109-112, 2011年

[学会発表](計14件)

(1) Tomohiro Nagata, Hiroki Mori, Takashi Nose, "Robust estimation of multiple-regression HMM parameters for dimension-based expressive dialogue speech synthesis," *INTERSPEECH 2013*, 2013/8/27, Lyon, France

(2) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "Statistical nonparametric speech synthesis using sparse Gaussian

processes," *INTERSPEECH 2013*, 2013/8/27, Lyon, France

(3) Takashi Nose, Misa Kanemoto, Tomoki Koriyama, Takao Kobayashi, "A style control technique for singing voice synthesis based on multiple-regression HSMM," *INTERSPEECH 2013*, 2013/8/26, Lyon, France

(4) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "Frame-level acoustic modeling based on Gaussian process regression for statistical nonparametric speech synthesis," *ICASSP 2013*, 2013/5/31, Vancouver, Canada

(5) Hiroki Kanagawa, Takashi Nose, Takao Kobayashi, "Speaker-independent style conversion for HMM-based expressive speech synthesis," *ICASSP 2013*, 2013/5/31, Vancouver, Canada

(6) Yu Maeno, Takashi Nose, Takao Kobayashi, Tomoki Koriyama, Yusuke Ijima, Hideharu Nakajima, Hideyuki Mizuno, Osamu Yoshioka, "HMM-based expressive speech synthesis based on phrase-level F0 context labeling," *ICASSP 2013*, 2013/5/31, Vancouver, Canada

(7) Vataya Chunwijitra, Takashi Nose, Takao Kobayashi, "A speech parameter generation algorithm using local variance for HMM-based speech synthesis," *INTERSPEECH 2012*, 2012/9/11, Portland, USA

(8) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "Discontinuous observation HMM for prosodic-event-based F0 generation," *INTERSPEECH 2012*, 2012/9/10, Portland, USA

(9) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "An F0 modeling technique based on prosodic events for spontaneous speech synthesis," *ICASSP 2012*, 2012/3/29, Kyoto, Japan

(10) Takashi Nose, Takao Kobayashi, "Recent development of HMM-based expressive speech synthesis and its applications," *APSIPA ASC 2011*, 2011/10/19, Xi'an, China

(11) Tomoki Koriyama, Takashi Nose, Takao Kobayashi, "On the use of extended context for HMM-based spontaneous conversational speech synthesis," *INTERSPEECH 2011*, 2011/8/30, Florence, Italy

(12) Tatsuhiko Saito, Takashi Nose, Takao Kobayashi, Yohei Okato, Akio Horii, "Performance prediction of speech recognition using average-voice-based speech synthesis," INTERSPEECH 2011, 2011/8/29, Florence, Italy

(13) Yu Maeno, Takashi Nose, Takao Kobayashi, Yusuke Ijima, Hideharu Nakajima, Hideyuki Mizuno, Osamu Yoshioka, "HMM-based emphatic speech synthesis using unsupervised context labeling," INTERSPEECH 2011, 2011/8/29, Florence, Italy

(14) Takashi Nose, Takao Kobayashi, "A perceptual expressivity modeling technique for speech synthesis based on multiple-regression HSMM," INTERSPEECH 2011, 2011/8/28, Florence, Italy

## 6 . 研究組織

### (1)研究代表者

能勢 隆 (NOSE TAKASHI)

東北大学・大学院工学研究科・講師

研究者番号：9 0 5 5 0 5 9 1