様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ５ 月 ２９ 日現在

機関番号：１１３０１
研究種目：挑戦的萌芽研究
研究期間：2016～2018
課題番号：１６Ｋ１３２５３

研究課題名（和文）平均声モーフィングを利用した日本語発音学習システムの研究開発

研究課題名（英文）Research and development of a Japanese pronunciation training system using average voice morphing

研究代表者
能勢　隆（NOSE, Takashi）

東北大学・工学研究科・准教授

研究者番号：９０５５０５９１

交付決定額（研究期間全体）：（直接経費）　　2,500,000円

研究成果の概要（和文）：本課題では、日本において非母語話者が日本語の発音学習を「低コストで」「手軽に」「確実に」行えるような新たな枠組の実現を目指した。具体的には複数の教師話者の音声により学習した平均教師声モデルによる統計的パラメトリック音声合成を利用し、音声の音韻や韻律（ピッチ・リズム）を特徴量毎に置換することで、従来よりも詳細で高精度な発音スコアのラベル付けを可能とした。この手法を用いて音韻、アクセント、リズムについて個別に発音スコアの予測モデルを学習し、非母語話者の発音スコアを予測することで、発音学習を効率的に行うことを実現した。

研究成果の学術的意義や社会的意義
法務省により公開されている在留外国人統計表によれば、日本における外国人の数は年々増加している。その一方で、英会話などに比べると発音の学習を提供するサービス、ソフトウェアは遥かに少ない。本課題で目指すシステムにより（１）非母語話者が発音の違いにより受ける社会的不利益やコミュニケーション力の低下などの問題が大幅に低減される。（２）構築した音韻・韻律別発音評定データベースを公開することで、多くの研究者に対してもより詳細な発音学習研究や応用が可能となる。などの社会的・学術的な波及効果が期待できる。

研究成果の概要（英文）：In this study, we aim to make a new framework of realizing low cost, convenient, and convincing system for a Japanese pronunciation training for non-native speakers in Japan. Specifically, we used a statistical parametric speech synthesis with an teacher average-voice model trained using multiple teachers' speech, and achieved a more precise labeling of pronunciation scores by using feature substitution technique for phonetic and prosodic parameters of speech. We trained a prediction model of pronunciation scores for phoneme, accent, and rhythm, and achieved an efficient pronunciation training method by predicting non-native speakers' pronunciation scores.

研究分野：音声合成、音声対話システム、音声認識、音声信号処理、音声情報処理

キーワード：e-ラーニング　コンピュータ学習支援（CALL）　発音学習　統計的パラメトリック音声合成　深層学習　韻律置換

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

１．研究開始当初の背景

外国語の学習をパソコンなどを用いて行うＣＡＬＬ(Computer-Assisted Language Learning)システムは、英会話学校などに比べ費用や場所といったコストが抑えられ、利用者が自主的・効率的に学習を行うことができるため、各種教育機関を中心に普及が進んでいる。一方で、日本ではその人口比から対象言語は英語が中心であり、比較的人口が少ない留学生などの外国人が非母語である日本語を学習するためのシステムは非常に限られているのが現状である。その中でも発音学習についてはテキストなどの教材により学習することが比較的困難であるため、低コストで学習を行うためには音声処理を導入した専用の CALL システムの活用が特に重要である。

これまでに英語の発音学習については音声データベースの構築[峯松'02]やシステムの開発[井本'01][阿部'01]を始め、商用のものも数多く存在するものの、学習者人口が少ない外国人のために公開されている日本語発音学習支援システムは三輪らによる「LESSON/J」や峯松らによるオンラインアクセント辞書「OJAD」によるプロジェクトなどごくわずかである。

２．研究の目的

本課題では、日本において非母語話者が日本語の発音学習を「低コストで」「手軽に」「確実に」行えるような新たな枠組の実現を目指す。具体的には複数の教師話者の音声により学習した平均教師声モデルによる音声合成とパラメータの置換を利用し、音韻・韻律別の発音評定データベースを構築し、発音スコア予測へと応用することを主たる目的とし、語学学習環境に恵まれない外国人を対象とした、日本語の総合的な CALL システムの開発のための萌芽的研究を行う。

３．研究の方法

研究期間は 3 年間であり、平成 28, 29 年度は発音スコアのフィードバックに焦点を当て、音韻・アクセント・リズムの個々の発音評価を厳密に行うために平均声と非母語話者データベース間の特徴量置換による評価刺激を作成し、主観評価実験により発音スコアデータベースを作成する。また最終年度にはこれらの機能を統合した発音学習システムを構築し評価実験を行う。具体的な項目について以下に述べる。

(1)非母語話者音声データベースの整備

研究期間を通して、学習者である日本語非母語話者の音声として音声資源コンソーシアムにて公開されている「留学生による読み上げ日本語音声データベース(JRF-DB)」を利用する。このコーパスには外国人留学生として男性 72 名、女性 69 名の音声が含まれている。研究期間を通して各音声の音素時間情報が必要となるため、この情報を音声認識で用いられる不特定話者モデルによる音素アラインメントにより推定し、手修正により正確な音韻・時間情報の付与を行う。

(2)平均声モデルの学習

発音の手本となる教師音声の生成には隠れマルコフモデル(HMM)に基づく音声合成[吉村'00]を用い、複数の教師話者（プロのアナウンサーやナレーター）の音声により学習した音響モデルを用いることで、個々の教師話者の癖を抑えたより平均的で望ましい合成音声が得られる。教師話者の音声としては音声合成において利用される ATR 日本語音声データベースに含まれる男性 6 名、あるいは女性 4 名の音声を用いることができるため、新たに収録を行う必要はない。

(3)音声パラメータの置換による実験用刺激の作成

音韻、アクセント、リズムのそれぞれの特徴について個別に発音スコアを付与するため、非母語話者音声データベースにおいて、音声分析によりスペクトル・基本周波数パラメータの抽出を行い、対象となる特徴以外については平均声のパラメータへと置換を行う。リズムについては平均声音声を生成する際に非母語話者音声の音素継続長を用いることにより反映させる。

(4)音韻・韻律発音スコアデータベースの構築

上記で作成した実験用刺激に対して複数の日本語話者による発音評価実験を行い、音韻、アクセント、リズムのそれぞれに対して発音スコアを付与したデータベースの構築を行う。この際、日本語話者としては発音訓練を受けた話者（アナウンサーやナレーター）とそのような訓練を受けていない一般話者の両方を対象とし、その相違についても分析する。

(5)発音スコアのモデル化と予測

構築した発音スコアデータベースを用いて、音韻・アクセント・リズム別に発音スコアのモデル化と予測を行う。モデル化にはサポートベクター回帰やニューラルネットワークなど、複数の手法を検討し、予測性能の比較を行う。予測のための特徴としては平均声音声とのスペクトル、対数基本周波数、音素継続長などの平均誤差・局所誤差・相関係数・それらの動的特徴量などを用いる。これにより、未知の学習者の音声に対しても発音スコアのフィードバックが可能となる。

４．研究成果

＜平成２８年度＞

音韻、アクセント、リズム毎の正確な発音スコアのラベル付けを行うための特徴量置換法を提案し、この手法を用いて非日本語母語話者の単語音声の主観評価を行い、有効性を示した。

具体的には、従来に比べ、ラベラー内、およびラベラー間でばらつきの少ない発音スコアが得ることができ、発音スコアデータベースを構築する際に有効であることがわかった。また、アクセントについて発音スコアデータベースを作成し、アクセントスコアの予測実験を行い、従来より高精度な予測スコアが得られることを示した。
＜平成２９年度＞
音声合成方式を従来の隠れマルコフモデルに基づく手法から深層学習に基づく手法へと変更し、より高品質な合成音声を生成することを実現した。また、実際に主観評価を行いアクセントとリズムについてのスコアを付与し、それに基づいてサポートベクター回帰に基づく予測実験を行ない、有効性を確認した。
＜平成３０年度＞
日本語音声合成において、非母語話者が聞きやすい音声の話速・ポーズ挿入位置について検討したものを論文としてまとめ公表するとともに、より聞きやすい音声を得ることを目的として、合成音声の自然性向上のため深層学習に基づいて差分特徴量に基づく制御法、および日本語アクセントの推定精度向上について取り組んだ。前者については、セグメント単位で平均化された基本周波数情報を用いて、従来の合成音声に対する差分となる特徴量を深層学習によりモデル化することで、自然性を保ちつつより柔軟な韻律の制御が可能となった。

５．主な発表論文等

〔雑誌論文〕（計 8 件）
①Analysis of Preferred Speaking Rate and Pause in Spoken Easy Japanese for Non-Native Listeners, Hafiyan Prafianto, Takashi Nose, Yuya Chiba, Akinori Ito, Acoustical Science and Technology, 査読有, vol. 39, 2018, pp. 92-100

②Segmental Pitch Control Using Speech Input Based on Differential Contexts and Features for Customizable Neural Speech Synthesis, Shinya Hanabusa, Takashi Nose, Akinori Ito, Proceeding of the Thirteenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 査読有, 2018, pp. 124-131

③Improvement of Accent Sandhi Rules Based on Japanese Accent Dictionaries, Hiroto Aoyama, Takashi Nose, Yuya Chiba, Akinori Ito, Proceeding of the Thirteenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 査読有, 2018, pp. 140-148

④Development and Evaluation of Julius-Compatible Interface for Kaldi ASR, Yusuke Yamada, Takashi Nose, Yuya Chiba, Akinori Ito and Takahiro Shinozaki, Proceeding of the Thirteenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 査読有, 2017, pp. 91-96

⑤Voice Conversion from Arbitrary Speakers Based on Deep Neural Networks with Adversarial Learning, Sou Miyamoto, Takashi Nose, Suzunosuke Ito, Harunori Koike, Yuya Chiba, Akinori Ito, Takahiro Shinozaki, Proceeding of the Thirteenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 査読有, 2017, pp.97-103

⑥韻律置換に基づく主観スコアを用いたアクセント自動評価の精度向上の検討，ハフィヤン・プラフィアント，能勢 隆，千葉祐弥，伊藤彰則，日本音響学会 2017 年春季研究発表会講演論文集，査読無，2017，pp. 251-252

⑦特徴量置換を用いた非日本語母語話者単語音声の主観評価，ハフィヤン・プラフィアント，能勢隆，千葉祐弥，伊藤彰則，日本音響学会2016年秋季研究発表会講演論文集，査読無，2017，pp. 249-250

⑧A Precise Evaluation Method of Prosodic Quality of Non-Native Speakers Using Average Voice and Prosody Substitution, Hafiyan Prafianto, Takashi Nose, Akinori Ito, Proceedings of the International Conference on Audio, Language and Image Processing, 査読有, 2016, pp. 208-212

〔学会発表〕（計 7 件）
①Segmental Pitch Control Using Speech Input Based on Differential Contexts and Features for Customizable Neural Speech Synthesis, Shinya Hanabusa, The Thirteenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2018

②Improvement of Accent Sandhi Rules Based on Japanese Accent Dictionaries, Hiroto Aoyama,

The Thirteenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2018

③特徴量置換を用いた非日本語母語話者単語音声の主観評価，ハフィヤン・プラフィアント，日本音響学会 2016 年秋季研究発表会，2017

④韻律置換に基づく主観スコアを用いたアクセント自動評価の精度向上の検討，ハフィヤン・プラフィアント，日本音響学会 2017 年春季研究発表会，2017

⑤Voice Conversion from Arbitrary Speakers Based on Deep Neural Networks with Adversarial Learning, Sou Miyamoto, The Thirteenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2017

⑥Development and Evaluation of Julius-Compatible Interface for Kaldi ASR, Yusuke Yamada, The Thirteenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2017

⑦A Precise Evaluation Method of Prosodic Quality of Non-Native Speakers Using Average Voice and Prosody Substitution, Hafiyan Prafianto, International Conference on Audio, Language and Image Processing, 2016

６．研究組織

(1)研究分担者

研究分担者氏名：千葉　祐弥

ローマ字氏名：CHIBA, yuya

所属研究機関名：東北大学

部局名：大学院工学研究科

職名：助教

研究者番号（8 桁）：30780936