

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 11 日現在

機関番号：12601

研究種目：挑戦的萌芽研究

研究期間：2012～2013

課題番号：24652115

研究課題名(和文)生成過程モデルと音声合成を用いた非母語音声の韻律体系化に基づく発音教育システム

研究課題名(英文)Pronunciation education system based on the systematization of non-mother tongue speech prosody using generation process model and speech synthesis

研究代表者

広瀬 啓吉(Hirose, Keikichi)

東京大学・情報理工学(系)研究科・教授

研究者番号：50111472

交付決定額(研究期間全体)：(直接経費) 2,800,000円、(間接経費) 840,000円

研究成果の概要(和文)：母語話者、学習者の基本周波数パターン分析を、生成過程モデルにより進め、フレーズ成分の特徴は、生起頻度に違いが見られるものの、基本的に言語間の差異が少ないことなどを示した。学習者の発音については、母語の影響が現れることなどを示した。学習者の非母語音声には、母語音声には見られないような基本周波数の動きが見られ、母語音声で学習した識別器は良好に動作しない。上昇下降の時点、傾斜を系統性に制御した合成音声を用いた知覚実験結果をもとに、高低判定を簡単な閾値判断によって行う手法等を開発した。また、HMM音声合成に生成過程モデルの制約を加えることを行った。日本語アクセント型発音教育システムを開発して評価した。

研究成果の概要(英文)：Fundamental frequency (F0) contours of speech by natives and learners are analyzed using the generation process model. Several findings, such as phrase components being less affected by language differences, are shown. As for utterances by learners, influence of their mother tongue is observed. Since learners utterances involve F0 movements not observable in natives utterances, accent type identifier trained using native utterances does not work well. To solve this problem, a series of perceptual experiments is conducted using synthetic speech with systematic control on F0 (points of F0 movements, slope coefficients). Based on the result, a threshold method of high-low decision of F0 is developed. Also, generation process model constraints are applied to HMM-based speech synthesis resulting in speech quality improvement. A pronunciation training system on Japanese accent type is developed and evaluated.

研究分野：人文学

科研費の分科・細目：外国語教育

キーワード：韻律体系化 非母語音声 生成過程モデル 音声合成 発音教育CALL 基本周波数パターン 単語アクセント 音声変換

1. 研究開始当初の背景

国際交流が急速に進んでおり、外国語教育の充実、特に、発音教育の充実が急務である。語学教師に頼った従来の教育では、量、質ともに不十分である。計算機を語学教育に活用する計算機援用語学教育 (CALL) に対する期待が高まり、既に、多くの発音教育用 CALL システムが開発されているが、その利用は拡大していない。教材や画面などに工夫を凝らしてはいるものの、教育内容が、教師音声に従った繰り返し発声を基本とするもので、発音教育で求められる、①発音誤りの具体的な指摘、②具体的な発音矯正方法の提示、③母語話者がどの程度理解するかを定量的な指標の提示、が行われないからである。この3条件を実現するためには、まず非母語音声の特徴を明確に定式化する必要があるが、このためには、従来の様に、母語話者と非母語話者の音声を分析して比較するだけでなく、音声合成を活用した知覚的な側面からの研究が必須となる。

2. 研究の目的

音声の韻律は、文内容の知覚に重要な役割を果たすとともに、その制御が適切になされていない音声は、対象母語話者にとって違和感が多く聞き辛いものとなる。また、意図・態度・感情・個性といったパラ・非言語情報の伝達にも重要な役割を果たしている。これまで、発音教育における韻律の重要性は十分認識されてはいるものの、有効な教育方策が確立されていないのが現状である。この大きな理由は、韻律は、語や句といった長時間に渡る動的な特徴であり、学習者の母語、習熟度が複雑に影響し合っ

て現れるため、音声の音響分析からだけでは、学習者の非母語音声の韻律を的確に把握し定式化することが困難なためである。

これに対し、代表者は従来から、対数基本周波数の時間変化パターン (F0 パター

ン) を全体的な起伏のフレーズ成分と局所的な起伏のアクセント/声調成分とに分離して表現する生成過程モデルを提案し、それを用いて、日本語のみならず諸言語の音声の F0 パターンの分析を、当該国研究者と連携を取って進めてきた。この結果、アクセント/声調成分は言語によって大きく異なるものの、フレーズ成分には言語共通の側面が多く見られることを明らかにし、また、学習者の日本語音声に見られる母語の影響を定量的に示した。生成過程モデルでは、言語情報、パラ・非言語情報と韻律との明確な対応が得られるため、HMM 音声合成や GMM 音声変換と組み合わせることで、発話機構の制約に基づいた適切かつ柔軟な韻律制御を音声合成で実現できる。本研究では、韻律の言語共通・固有の側面、韻律に現れる学習者母語の影響、に対する理解に基づき、生成過程モデルの枠組みで系統的な韻律制御を行って音声合成し、聴取実験によって評価することで、非母語話者音声の韻律を定式化し、発音教育で前記①～③を実現することを目的とする。これによって、特に韻律の側面から、従来になく効果的な発音教育システムを達成する。

3. 研究の方法

主として、日本語、中国語、英語を対象として研究を進める。この3言語を選択する理由は、これらが、アクセント、声調、ストレスに代表されるように、それぞれ大きく異なる韻律的特徴を有し、韻律の異同を調べるには最適の組合せであると共に、発音教育の面から重要な言語であるからである。具体的には下記のように研究を進める。

(1) 韻律比較研究をするために適した文を各言語について選定し、母語話者について音声を収録する。また、それぞれの言語の組み合わせで、初級、中級それぞれの学習レベルの非母語音声を収録する。同時に学

習者の母語音声も収録する。平静の朗読音声のみではなく、意図・態度・感情といったパラ・非言語情報を制御した音声も収録する。得られた音声資料について、音素ラベリングを行うとともに、生成過程モデルの指令パラメータ抽出とそれによる韻律ラベル付けを行って韻律コーパスを作成する。代表者が開発した指令パラメータ自動抽出を活用して作業の効率化を図る。

- (2) 韻律の分析結果をもとに、生成過程モデルに現れる各言語の韻律的特徴を、語義、統語、談話(焦点)などの言語情報、意図・態度・感情といったパラ／非言語情報との関係といった観点を含めて定量的に比較する。得られた結果に基づき、韻律における言語共通の側面と言語固有の側面を明らかにした上で、学習者母語音声の韻律が非母語音声の韻律に与える影響を明確にする。また、習熟度によって、この影響がどの様に変化するかを調べる。得られた結果から、非母語音声の韻律の特徴の基本的な定式化を行う。
- (3) 非母語音声の特徴づける個々の韻律的特徴を整理した上で、TD-PSOLA 等によりそれらを系統的に変化させた音声を合成し、学習対象言語の母語話者による聴取実験を行う。これによって、学習者の母語、学習レベルを考慮に入れた上での非母語音声の韻律の特徴の詳細な定式化を達成する。
- (4) HMM (hidden Markov model) 音声合成、GMM (Gaussian mixture model) 音声変換において、生成過程モデルの枠組みで基本周波数パターン生成を行う音声合成システムを、韻律の観点から改良を加えるとともに、話者変換、言語変換の枠組みを実装する。
- (5) 学習者の非母語音声について、個々の韻律的特徴の発音誤りを検出した上で、母語話者にどの程度許容されるかを評定

することを行う。母語話者(教師)に評価結果を示し、その妥当性を検証する。この評定によって、学習者は自身のレベルを具体的に知ることが出来、学習に対するモチベーションを上げ、さらには学習達成時期を判断することが出来る。

- (6) 生成過程モデルによる韻律変換と HMM 音声合成・GMM 音声変換を統合した合成システムを用い、詳細な特徴の制御を行って音声を変換・合成し、聴取実験を行って非母語音声の韻律の特徴の解析を進める。
- (7) ある話者の母語音声から、話者の声質を保った上で、学習対象言語の音声を合成する言語変換を目指す。そのために、変換の参照となるバイリンガル音声を複数の話者について用意し、分節的特徴と韻律的特徴を、それぞれの言語の音声の特徴空間内での位置として記述する。
- (8) 学習者の非母語音声の発音誤りを修正し、言語変換により対象言語として理想的な発音を合成することを目指す。
- (9) 学習者の非母語音声の発音誤りを指摘し、母語話者による許容度を示した上で、個々の誤りを修正して再合成し、発音矯正として提示する発音教育システムを構築する。実際に試用して、学習効果を確認する。

なお、研究の規模から、上記 6, 7, 8 については、予備実験に止める。

4. 研究成果

母語話者の発話の韻律の言語間比較とともに、特に日本語の韻律に関する発音教育を中心として研究を進めた。まず、日本語、英語、中国語の母語音声コーパスを、海外研究者の助力を得て取得し、その基本周波数パターンの分析を、生成過程モデルにより進めた。その結果、英語については、従来、音源強度、音素持続時間との関連が強いとされる強勢、

弱勢の表現に、F0 パターンが大きく関係し、生成過程モデルにより強勢、弱勢が良好に区別されること、中国語については、各声調が、正、負あるいは両者の結合した指令により良好に表現され、談話の焦点は指令の絶対値により表わされることなどを明らかにした。また、統語といった、アクセント型等と比べて広い範囲に渡るフレーズの特徴は、指令の生起頻度に違いが見られるものの、基本的に言語間の差異が少ないことを明らかにした。なお、これと関係して、観測される F0 パターンに対し、生成過程モデルによる分析（指令の抽出）を自動で行う新しい手法を開発した。従来の手法と異なり、この手法では、観測された基本周波数（F0）から平滑化と補間によって連続的なパターンを求めることはせずに、母音の F0 パターンの高低に着目する。これと、アクセント型情報からフレーズ成分の概形を求め、指令を求めた後、アクセント指令を求める。アクセント型に誤りのある発声に対しては、性能の低下があるが、正しい発声に対しては、従来手法と比べ、抽出性能に格段の向上がみられた。この手法は、エキスパートが指令抽出を行う際の過程をシミュレートしたものと言える。ただし、アクセント型規則といった日本語の知識を導入しているため、そのままでは他言語への適用は困難であり、引き続き研究が必要である。

非母語話者（学習者）の発音については、日本語発音を収集し、単語アクセント型、長母音、促音などの特殊拍音素における発音誤りを調査した。その結果、学習者の母語の影響が F0 パターンに現れること、特殊拍音素は対応する通常音素の発音に影響を与えることなどを示した。例えば日本語文を母語話者が発声した場合と中国語を母語とする日本語学習者が発声した場合について比較すると、母語話者では、F0 パターンによる語句のグルーピングが良好に行われているのに対し、学習者音声では、個々の語句がばらば

らに発声されている。生成過程モデルによる分析を行い、アクセント指令の数に明確な違いがみられた。学習者の非母語音声では指令の数が多く、単語単位でのアクセント型の生成、アクセント結合に問題があることが分かる。なお、英語を母語とする日本語学習者の音声の分析も行い、その F0 パターンには中国語を母語とする学習者との違い、すなわち母語の影響がみられた。

標準中国語音声については、母語音声、非母語音声の韻律の特徴比較を行い、非母語音声のリズムパターンに、英語に代表される **Stress-timed** 言語の特徴が表れている様に見えるものの、単に学習者の不慣れから来るものであって、詳細にみると母語音声の影響が表れることを示した。

学習者の音声では、母語話者の音声では見られないような F0 の動きが見られる。例えば、標準日本語の単語アクセントでは、母語音声には見られない上昇下降を繰り返すパターンが見られる。このため、母語音声を用いて学習したアクセント型識別器では、学習者の非母語音声のアクセント型の誤りを的確に指摘することは出来ない。そこで、アクセント型の識別において、2 モーラ単位での F0 パターンの上昇下降が重要なことを考慮し、上昇下降の時点、傾斜を系統性に制御した合成音声を用いて、母語話者による高低の知覚との関係を調べた。それをもとに、高低の判定を簡単な閾値判断によって行う手法を開発し、統計的な **SVM** (support vector machine) による手法を超える性能を達成した。また、学習者音声のアクセント型が母語話者にどの程度正しく知覚されるかを数値として示すことを行った。

日本語の特殊拍音素は時間長制御と関係し、非母語話者には発音習得が難しい音素として知られている。特に促音が含まれると知覚されるか否かについて、種々の先行・後続母音の文脈で、発話速度を制御して合成した

無意味単語の聴取実験を行い、時間長に現れる判断境界を詳細に調べた。得られた結果をもとに簡単な閾値判断によって促音の識別を行う手法を開発し、HMMを用いる手法を大きく超える性能を得た。

音声合成については、HMM音声合成で、生成過程モデルの制約を加える2手法を開発した。1つ目は、生成過程モデルにより近似されたF0パターンをHMM学習に用いるもので、2つ目はHMMから生成されたF0パターンを生成過程モデルで近似するものである。両者とも、従来のHMM音声合成と比較して、合成音声品質の向上を達成した。この手法では、生成過程モデルでF0パターンが表現されるため、モデルの指令を操作することで、焦点などの高次言語情報の付与、スタイル・話者の変換を行うことが可能である。また、前述のようにF0パターンを系統的に変化させて、アクセント型知覚などの判断境界を調べるにも有用である。焦点付与、声質変換を行って手法の有用性を確認した。なお、声質変換の際の分節的特徴の取り扱いについて、話者の特徴を行列変量で表す変換手法を開発し、その有効性を確かめた。この手法は、多数話者の音声を同時に扱うことが可能で、研究方法7,8の基盤となるものである。

得られた成果をもとに、学習者が発声した音声のF0パターンから、アクセント型が正しく発声されているかを判定し、自動的に修正して、学習者の声質の音声として提示する、日本語アクセント型発音教育システムを開発した。

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6件)

- ① Greg Short, Keikichi Hirose, and Nobuaki Minematsu, "Automatic recognition of gemination in Japanese motivated by perceptual experiments," *Acoustical Science and Technology*, Vol.35, No.2, pp.73-85 (2014-3). 査読有
<https://www.jstage.jst.go.jp/browse/ast>

- ② Keikichi Hirose, "Toward flexible and systematic control of fundamental frequencies in HMM-based speech synthesis," *Journal of EPSJ (English Phonetics Society of Japan)*, Vol.18, pp.121-128 (2013-12). 招待論文、査読無

<http://www.cc.kochi-u.ac.jp/~tamasaki/EPSJ.htm>

- ③ Greg Short, Keikichi Hirose, and Nobuaki Minematsu, "Japanese lexical accent recognition for a CALL system by deriving classification equations with perceptual experiments," *Speech Communication*, Vol.55, Issue 10, pp.1064-1080 (2013-11/12). 査読有
<http://www.sciencedirect.com/science/article/pii/S0167639313000927>

- ④ Orphan Krityakien, Keikichi Hirose, and Nobuaki Minematsu, "Generation of fundamental frequency contours for Thai speech using the tone nucleus model," *Journal of Signal Processing, Research Institute of Signal Processing*, vol.16, no.4, pp. 135-138 (2013-7). 査読有

<https://www.jstage.jst.go.jp/browse/jsp>

- ⑤ Qinghua Sun, Keikichi Hirose, and Nobuaki Minematsu, "A method for generation of Mandarin F_0 contours based on tone nucleus model and superpositional model," *Speech Communication*, Vol.54, Issue 8, pp.932-945 (2012-10). 査読有

<http://www.sciencedirect.com/science/article/pii/S0167639312000349>

- ⑥ Tatsuya Matsuda, Keikichi Hirose, and Nobuaki Minematsu, "Applying generation process model constraint to fundamental frequency contours generated by hidden-Markov-model-based speech synthesis," *Acoustical Science and Technology*, Acoustical Society of Japan, Vol.33, No.4, pp.221-228 (2012-7). 査読有

<https://www.jstage.jst.go.jp/browse/ast>

[学会発表] (計19件)

総て論文集有り。④～⑦、⑩、⑫、⑬以外は査読有り。

- ① Ya Li, Jianhua Tao, Keikichi Hirose, Wei Lai, and Xiaoying Xu, "Hierarchical stress generation with Fujisaki model in expressive speech synthesis," *Proceedings of International Conference on Speech Prosody*, Dublin, May 20-23, pp.1032-1036 (2014-5).

- ② Tomoyuki Mizukami, Hiroya Hashimoto, Keikichi Hirose, Daisuke Saito, and Nobuaki Minematsu, "Selection of training data for HMM-based speech synthesis from prosodic features - Use of generation process model of fundamental frequency contours -," *Proceedings of International Conference on Speech Prosody*, Dublin, May 20-23, pp.1042-1046 (2014-5).

- ③ Wentao Gu and Keikichi Hirose, "Rhythmic patterns in Native and non-native Mandarin Speech," *Proceedings of International Conference on Speech Prosody*, Dublin, May 20-23, pp.592-596 (2014-5).
- ④ 百武恭汰, 橋本浩弥, 齋藤大輔, 峯松信明, 広瀬啓吉, "生成過程モデルにおける F0 パターン差分を考慮した HMM 音声合成の実験的検討," 日本音響学会全国大会講演論文集, 日本大学, 3 月 10 日-12 日, 1-R5-14, pp.407-408 (2014-3).
- ⑤ Wentao Gu and Keikichi Hirose, "Rhythmic Patterns of Nonnative Mandarin Speech," 日本音響学会全国大会講演論文集, 日本大学, 3 月 10 日-12 日, 3-6-16, pp.357-360 (2014-3).
- ⑥ 土井秀信, 齋藤大輔, 峯松信明, 広瀬啓吉, "行列変数ガウス混合分布に基づく声質変換の検討," 日本音響学会全国大会講演論文集, 日本大学, 3 月 10 日-12 日, 1-R5-21, pp.425-428 (2014-3).
- ⑦ 橋本浩弥, 広瀬啓吉, 峯松信明, "文節を基本単位とした基本周波数パターン生成過程モデルのパラメータ自動抽出," 日本音響学会全国大会講演論文集, 豊橋技術科学大学, 9 月 25 日-27 日, 1-P-5a, pp.327-328 (2013-9).
- ⑧ Hiroya Hashimoto, Keikichi Hirose, and Nobuaki Minematsu, "Context labels based on "bunsetsu" for HMM-based speech synthesis of Japanese," *Proceedings 8th ISCA Workshop on Speech Synthesis (SSW-8)*, Barcelona, August 31-September 3, pp.35-39 (2013-8-31).
- ⑨ Oraphan Krityakien, Nobuaki Minematsu, and Keikichi Hirose, "Generation of fundamental frequency contours for Thai speech synthesis using tone nucleus model," *Proceedings INTERSPEECH 2013*, Lyon, August 26-29, pp.1037-1041 (2013-8).
- ⑩ 川口拓也, 橋本浩弥, 広瀬啓吉, 峯松信明, "基本周波数パターン生成過程モデルの指令差分に基づく焦点制御の改良," 日本音響学会全国大会講演論文集, 東京工科大学, 3 月 13 日-15 日, 3-P-34b, pp.505-506 (2013-3).
- ⑪ Keikichi Hirose, Hiroya Hashimoto, Jun Ikeshima, and Nobuaki Minematsu, "Use of generation process model for synthesizing fundamental frequency contours in HMM-based speech synthesis," *Proceedings IEEE International Conference on Signal Processing (ICSP'12)*, Beijing, October 22-24, pp.575-578 (2012-10).
- ⑫ ショート・グレッグ, 広瀬啓吉, 峯松信明, "アクセント核の知覚と母音長との関係の基礎的検討," 日本音響学会全国大会講演論文集, 信州大学, 9 月 19 日 - 21 日, 3-Q-31, pp.429-430 (2012-9).
- ⑬ Krityakien Oraphan, Hirose Keikichi, and Nobuaki Minematsu, "Tone nucleus model for Thai language speech synthesis," 日本音響学会全国大会講演論文集, 信州大学, 9 月 19 日 - 21 日, 3-Q14 pp.391-392 (2012-9).
- ⑭ Raymond W. M. Ng, Thomas Hain, and Keikichi Hirose, "An alignment matching method to explore pseudosyllable properties across different corpora," *Proceedings INTERSPEECH 2012, Portland*, September 9-13, 4 pages in CD-ROM Proceedings (2012-9-11).
- ⑮ Daisuke Saito, Nobuaki Minematsu, and Keikichi Hirose, "Effects of speaker adaptive training on tensor-based arbitrary speaker conversion," *Proceedings INTERSPEECH 2012, Portland*, September 9-13, 4 pages in CD-ROM Proceedings (2012-9-10).
- ⑯ Hiroya Hashimoto, Keikichi Hirose, and Nobuaki Minematsu, "Improved automatic extraction of generation process model commands and its use for generating fundamental frequency contours for training HMM-based speech synthesis," *Proceedings INTERSPEECH 2012, Portland*, September 9-13, 4 pages in CD-ROM Proceedings (2012-9-10).
- ⑰ Keikichi Hirose, Hiroya Hashimoto, Jun Ikeshima, and Nobuaki Minematsu, "Fundamental frequency contour reshaping in HMM-based speech synthesis and realization of prosodic focus using generation process model," *Proceedings of International Conference on Speech Prosody*, Shanghai, May 22-25, SS1-3, pp.171-174 (2012-5).
- ⑱ Raymond W. M. Ng and Keikichi Hirose, "Automatic segmentation of English words using phonotactic and syllable information," *Proceedings of International Conference on Speech Prosody*, Shanghai, May 22-25, PS1A-6, pp.27-30 (2012-5).
- ⑲ Shuhei Kato, Greg Short, Nobuaki Minematsu, and Keikichi Hirose, "Effects of learners' language transfer on native listeners' evaluation of the prosodic naturalness of Japanese words," *Proceedings of International Conference on Speech Prosody*, Shanghai, OS2-2, May 22-25, pp.198-201 (2012-5).

6. 研究組織

(1) 研究代表者

広瀬 啓吉 (HIROSE, Keikichi)
 東京大学・大学院情報理工学系研究科・教授
 研究者番号 : 50111472

(2) 研究分担者

なし

(3) 連携研究者

河合 剛 (KAWAI, Goh)
 北海道大学・外国語教育センター・准教授
 研究者番号 : 70312981