

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 4 日現在

機関番号：12102

研究種目：基盤研究(B)

研究期間：2012～2014

課題番号：24320091

研究課題名(和文) 音声認識技術を取り入れた日本語発音自学システムの作成と試用

研究課題名(英文) The development and trial of software to self-study Japanese prosody using an automatic speech recognition system

研究代表者

松崎 寛 (MATSUZAKI, Hiroshi)

筑波大学・人文社会系・准教授

研究者番号：10250648

交付決定額(研究期間全体)：(直接経費) 14,000,000円

研究成果の概要(和文)：発音学習において、学習者に音響学的な波形やピッチ曲線を見せ、そこから適切な情報を読み取らせる従来の方法は、困難を伴う。そこで本研究では、音声認識技術を活用し、入力音声のリズムを視覚的に補正した「プロソディグラフ」を自動出力する新たなシステムを開発した。しかし、ピッチ曲線における高さをそのままのタイミングで示すと、アクセントの下がり目が遅れて生じる「おそさがり」現象により学習が阻害される。そこで、ピッチ情報を1拍分前にずらす修正方法を適用し、表示した結果、ほとんどの例文が妥当な表示となった。しかし、文末イントネーションの表示方法には、未解決の課題が残った。

研究成果の概要(英文)：This project aims at developing a new system that outputs a visual representation of prosody. It employs an automatic speech-recognition system and analyzes the segmental and prosodic features of a speech to provide an easy-to-understand visual feedback to Japanese learners. The problem of delayed F0 fall is corrected by assigning the F0 value to the mora that immediately precedes the segment in which the pitch drops. However, the inaccuracy in representing sentence-final intonation remains unsolved.

研究分野：日本語教育

キーワード：外国語 教育学 教育工学 発音学習 音声合成 音声学 音声認識

1. 研究開始当初の背景

日本語音声教育をとりまく環境は、近年、大きく変わりつつある。これまで、ミニマル・ペア練習を中心とした正確さ重視の発音指導方針は、コミュニケーション・アプローチの台頭とともに、「通じれば良い」と捉えられ、一部では音声教育不要論が語られることさえあった。しかし、ある研究者が様々な職種の人々に聞き取り調査を行った結果では、「ビジネスをするなら、きちんとした発音で話すべき」「医療の現場では、単なる会話では済まされない」と考える日本人も、少なからず存在することが明らかになっている。外国人も「公的な場面」においては、流暢で正確な発音を要求されるのである。

特に、ビジネスにおける「高度人材」、外国人看護師や介護福祉士は、正確な意思疎通が要求されるとともに、従来の第二次産業的労働力としての外国人とは異なるパーソナルコミュニケーション、すなわち、サービス業としての「失礼にならない話し方」=「感情的評価を下げない発音」が要求される職種でもある。

本研究は、そのような時代の要請を受け、新たな発音学習のための、自学用コースウェアの開発を行うことを目的としている。教育の現場においては、発音のために正規授業時間の多くを割くことは難しい状況にある。時間的な問題以外にも、教師に正確なモデル発音能力や音声教育に関する知識がない、何をどこまで教えたらよいかという明確な基準がない、学習意欲を高める練習方法の手持ち札が少ない等のさまざまな問題があり、教室での発音指導実施が困難となっていることが明らかになっている。

「発音の自学自習」に関して、学習者に音源を渡し、各自練習するよう指示したとしても、学習者が妥当な発音基準を持って自らの発音を適切に修正することには、あまり期待できない。すなわち教師が学習者の発音を評価し、修正する以外に適切な方法がない。従来の日本語 CALL 教材における、コンピュータが学習者の発音を判定する技術も、精度の問題や個人差の問題が大きく、実用化に至っていなかった。以前の音声認識エンジンは、話し手の癖を覚えさせるための例文録音（エンロールメント）を長時間に渡って行わなければ、認識率が非常に低くなるという問題があり、外国人学習者の誤った発音も適切に処理できなかった。

しかし近年の技術革新により、不特定多数の話者の発話を処理できる音声認識エンジンが数多く開発され、注目されている。たとえば、英語教育の分野では、産業技術総合研究所が IT 企業「プロンテスト」と共同開発をした発音学習用のソフトが市販され、2011 年度から茨城県の公立中に導入されている。

日本語発音学習ソフトとしては、アドバンス・メディア社と長沼スクールが共同開発した「AmiVoice」を取り入れた学習ソフト

「AmiVoice CALL Web -Japanese-」がある。これは 2009 年度より長沼スクールに導入された子音・母音・拍感覚の練習用ソフトで、製品化もされている。

このソフトは、教師の指導の下に教室で使用されることを想定して開発されたものであるが、これを自学自習用教材とするには、更なる改良が必要である。たとえば、学習者の発音の韻律情報については、画面のモデル音と自分の声の波形表示およびピッチ曲線でしか把握することができない。高さや音色の違いを確認させるといふ点では、この波形表示とピッチ曲線の対応を自分で見出して情報を統合し、学習者が音の高低や声帯振動の情報を読み取ることは、難しいと思われる。

そこで、本研究代表者は、基盤研究(B)「音声認識技術を取り入れた日本語発音自学システムの開発」(課題 ID 21320094)で共同研究を行い、入力音声のリズムを補正した「プロソディグラフ(高さ・長さを音符のように示した視覚教材)」を自動的に出力する「Prosody Tuner」を開発した。2011 年度の開発画面(一部)は、図 1 のようである。

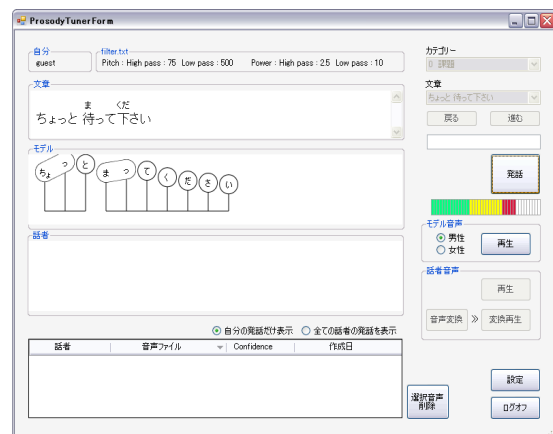


図 1 「Prosody Tuner2011」の画面

このソフトの機能を箇条書きにすると、以下のようになる。

- (a)モデル音声および入力された学習者の発音を、子音・母音で自動的に分節する
 - (b)それぞれの音節(短い円・破線の円・長い円)を入力された高さにあわせ適切な箇所に表示
 - (c)上段にモデル音、下段に学習者発音が表示され、ボタンクリックで音声がかかる。視覚的・聴覚的比較が可能
 - (d)学習者の発音は、円を上下に操作することにより高さの合成・再生が可能
 - (e)モデル音声の性別を選択できる。データとプロソディ情報は、話者毎に保存可能
- (d)のように、プロソディグラフの各円の高さを変えることで、話者はモデル発音と自分の発音の違いを視覚的に確認できるとともに、モデル音に近く合成された音声を聴覚的に確認することで、到達目標をより明確に把握することができる。しかし学習者への試用

の結果、この方法では妥当な合成音声を得られにくいことから、3文のみではあるが、(f)「韻律一発変換ボタン」によるモデル音の韻律と学習者発音の高さ合成が行えるようにした。また、(g)「高さ調節スライダー」による男女の声の高さの違い等の手動修正 (h)「日本語学習者誤用データベース」を元に、有声・無声、/ツ/・/チュ/、母音の長短など、39のエラーを指摘できる機能も付し、その他、文字の位置を円内から下部に移動し、高さの変化を見やすく表示した。

2. 研究の目的

以上のような背景から、本研究では、音響音声学的技術を取り込んだ教材を開発するにあたり、より効果的な学習が起こるにはどうすれば良いのかを追究することを目的とする。具体的には、音声学、教育方法学、教育工学、認知言語学、教育心理学、第二言語習得論等の様々な視点から基礎的・応用的調査研究を行うとともに、それらの成果を踏まえて開発された教材を試用する実践研究を行い、フィードバック結果を分析して、教材開発に役立てるものである。

研究目的を達成するために、次のような計画・方法で進めていく。

(1) 学習内容の深化：学習者の誤用傾向や母語特性にあわせた音配列の選定、学習者の日本語能力にあわせた例語・例文の選定。音声データベースの構築。大量のコンテンツ作成。

(2) 学習方法の追究：自律学習を可能にするコースウェアの開発。韻律表示方法の改善。発音モニター能力を向上させる練習方法の開発。効果的フィードバックおよび学習者の発音能力の伸長を記録・参照するためのシステムの開発。音声合成技術を援用したモデル発音提示方法の工夫。発音評価方法の妥当性・信頼性を高めるための基礎研究。

(3) 教材評価：国内外の日本語教育関係者の協力を得て、教材を試用してもらう。そのフィードバックをもとに教材を改良。

3. 研究の方法

2012年度前半は、上記のソフトをさらに改良することから始め、その試用成果の一部を、日本語教育国際研究大会で発表した。

しかし、「Prosody Tuner」に更なる改良を加えるには、技術的な問題や多額の開発費用、そして継続的なAmi Voiceのライセンス使用料が生じることなどがわかった。

そこで、最終的にWEB上で無料公開を行うために、同等の機能を持つCALL教材を、オープンソースのフリーソフトを活用して作成し、常に改良を加えていける体制を整えることに研究方針を変更した。

2012年度後半は、ひとまず、下記の機能を備えたWEBシステムのひな形を完成させた。

(1) 学習者による文読み上げ音声の韻律を「プロソディグラフ」で示す

(2) 学習者の誤った韻律を、モデル音の韻律と合成して正しく補正し、学習者に聞かせる (1)の基本周波数推定プログラムには「DIO」を、(2)の韻律補正には「WORLD」(森勢将雅氏開発のフリーソフト <http://www.slp.is.ritsumei.ac.jp/~morise/world/>)を使用したが、基本周波数推定や、音声合成のコア部分に手をつけることは困難を極め、学習者の最適な発音を誘導するプロソディグラフの描画方法や、録音時のノイズへの対応策などに困難を生じた。

そこで大学内外の、音声認識・音声合成専門の研究者を訪問し、専門的見地からアドバイスを受けるとともに、開発に携わる複数の研究協力者の確保に向けて行動した。

2013年度は、システム改良やプログラミングを行っていた研究協力者の大学院生が就職のため研究組織を離れるという更なる問題が生じたため、音声認識および教材開発に明るい業者の選定および折衝に、殊の外、時間を費やした。このソフトは最終版においても録音判定、音声合成、表示方法への対応に大きな問題が残ったため、開発したソフトを学習者に試用する段階まで至ることができなかった。

そこで、2014年度は、再度計画を見直し、「Prosody Tuner 2011」を作成した業者に再開発を依頼して、公開方法に関する交渉を行った。その結果、「学術目的に限っての3年間の無償公開」が行える許可が最終的に得られた。ダウンロードは、以下より。

<http://www.u.tsukuba.ac.jp/~matsuzaki.hiroshi.fp/>

2014年度の開発における最大の課題は、「音の高さの表示が妥当でない箇所が、発音学習の妨げになっている問題」をどう解決するかであった。たとえば、図1でいえば、「ちょっと」「待って」等は、音韻的には本来、語頭が一番高いはずだが、実際には次拍のピッチが高く表示される。これは、「おそ下がり」と呼ばれる現象で、物理的なピッチ下降が、実際の聴覚印象よりも遅れて生じる現象である。

音響分析の結果と内省の結果が大きく異なる現象としては、「自然減衰」と「おそ下がり」が知られているが、学習者のソフト試用において、自然減衰による句全体の下降がアクセント核での下降と混同される「はや下がり」の可能性のありそうな箇所は、教師が横について「どこで下がっているように見えるか」を確認し、また学習者が有していた「アクセントに関する知識」を確認した結果、この点が学習の妨げとなる事例は見られなかった。

しかし、アクセントの下がり目のずれに関しては、同様の方法で「どう見えるか」と「アクセント知識」を確認した結果、表示が学習の妨げとなる例が、多く見られた。そのため、

本研究者が横について、画面表示と下がり目のずれに関して一語一語解説を加えながら学習を進めざるをえなかった。すなわち、誤った表示が学習者の自学を阻害する原因となっているといえる。

この問題を解決するには、いくつかの改善方法が考えられる。まず、もっとも直接的なものは、「アクセントに関する辞書情報をモデル発音に適用して表示を補正する方法」である。しかしそれでは、学習者の発音をプロソディグラフ化して示す際に辞書情報が利用できないため、学習者の誤り発音を、モデル音と同じ基準で比べることができなくなる。つまり、モデル発音と学習者の発音に同様の方法で補正を加える方法を開発しなければならない。

そこで、「ピッチ情報を1拍分前倒して適用すると、どのくらい表示の問題が解決するか」を、システム収録の204文530句における男声のみを対象として試算した。

4. 研究成果

調査結果を表1に示す。図2の場合、/カド//オロ シテ/は改善する一方、/テマエデ/は/オロ シテ/と類似してしまうが、全体的な自然減衰とあわせれば、効果としては/カド/の改善の効果が大きく、改悪ではない。

表1 1拍ずらして改善/改悪された句数

a. 句数	b. 有核	c.おそ下がり	d. 改善	e. 改悪
530	374	234	200	41

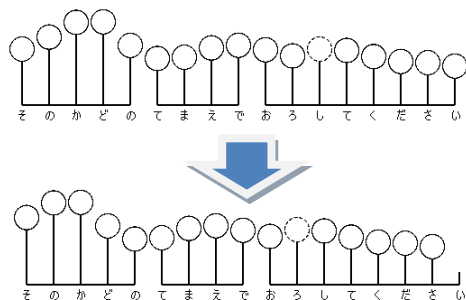


図2 おそ下がり補正イメージ

「b.有核」語の6割に「c.おそ下がり」が見られ、その85%が1拍ずらして「d.改善」する。残り1割は、2拍以上ずらさないといえないものである。正しいものがずれてしまう「e.改悪」41を引くと、総計68%が1拍ずらして改善されたことになる。

ピッチ情報を1拍分前倒した分、欠けることとなる末尾1拍は、それまでのピッチ推移から高さを予測して処理することで対応したが、これは逆に、上昇調への改悪箇所も生じることとなった。この表示の妥当な処理方法の追究については、今後の課題である。

以上、前述の Prosod

y Tuner に搭載した(a)~(h)の機能に加えて、「Prosody Tuner 2015」では、以下の機能を開発・登載

(i)「おそ下がり」を修正して表示が聴覚印象に近づくように改善

(j)一発変換によるモデル韻律との合成が全ての文で行えるようにした

(k)韻律の逸脱度を拍ごとに S,A,B,C で評価する機能を加えた

(k)に関しては、初期版では文全体に対して1つのスコアが与えられる評価システムをつけたのだが、学習者から「どこが間違っているのかに関するフィードバックがほしい」とのコメントがあったため、拍ごとの逸脱度を示す方式に変更した。

ただし、この評価の精度には、まだかなりの問題があるため、今後、ラベリング作業等、日本人発音評価研究をさらに深め、その成果を評価機能に搭載する必要がある。

その他、画面拡大に伴うゲージ移動の不具合および、画面切り替えに伴う評価表示の消滅の不具合を修正し、男女の声の高さに合わせて表示を最適化するようにした。また、音声の合成と出力を1クリックで行えるようにした。これらの機能修正のアイデアは、すべて学習者の試用時のコメントから得られたもので、最終的には、かなり使い勝手がよくなったと言える。



図3 「Prosody Tuner 2015」の画面

2015年2月に、広島大学留学生6名(中国3,台湾1,トルコ1,ペルシア1)の日本語教育専攻大学院生/研究生)に教材を試用させ、反応を観察・撮影した。

課題文は、高さ表示が比較的妥当と判断された次の6文である。

- ・じゃ、お待ちしています
- ・歩いてどのくらいかかりますか
- ・その角の手前でおろしてください
- ・どうして来なかったんですか
- ・中古品って人が使った物っていう意味ですね
- ・この問題が分からないんですけど教えてい

ただけませんか

試用後、使用感や要望に関する半構造的インタビューを行った。内容は、音声学や発音学習経験、ソフト試用の感想、改善点などについてで、時間は1時間前後であった。

まずモデル音を聞く前に例文を読ませ、その後、モデル音との比較や、自分の発音の表示とモデル音の表示の比較を通して練習させた。調査者の介入は極力避けたが、協力者が誤用箇所に気付けない場合、「この辺を比べてみて」「どちらが高くなっていますか」等の質問をした。

【試行中の行動】

多くの協力者は普段の読み方を最初に録音し、モデル音と自分のプロソディグラフを見比べ、違う箇所をモデルに近づけようと練習を繰り返した。より積極的に活用しようとした協力者は、最初の音読の前にモデル音の表示を見ながら小声で何度もリハーサルした。練習でも高さに合わせて手を動かすなどの行動が見られた。一方、わずかな傾斜のズレにこだわり、もっと問題の大きい他の誤用に気付けないことも多かった。本人が自覚する苦手な発音に意識が焦点化されやすいためかもしれない。

合成音声は、今回はレベル的に必要のある者がおらず、好意的な評価が得られなかった。表示やプロソディ評価(S,A,B,C)はまだ精度が低く、妥当な表示/評価にならないことが多々あり「どうして(モデル音と自分の発音の表示が)同じにならないか」「(発音の評価が違うが)さっきと今と何が違いましたか」という戸惑いも見られた。

【インタビュー】

試用に関しては概ね好評で、面白い、楽しい、使いたいという感想が聞かれ、「自分の発音のプロソディグラフがすぐ見られる」「何度も繰り返し録音して聞ける」「その都度評価が出る」が良い点として挙げられた。

「推測のためのルールが知りたい」という意見もあった。今回は無作為に例文を練習させたが、将来的には、文イントネーションやアクセントの学習を体系的に行うカリキュラムを整え、メイン画面から練習画面に移動して学習させる方法を考える必要がある。

今後も国内外の日本語教育関係者の協力を得て、試用を続け、さらなる教材の改良を行うとともに、ケースを蓄積してCALL教材使用時の学習者の学習行動の分析を深めていきたい。

5. 主な発表論文等

〔雑誌論文〕(計 2件)

松崎寛・高橋恵利子「おそ下がりに対応したプロソディグラフ自動出力システムの開発」『日本語教育方法研究会誌』査読無, Vol.22, No.1, 2015, pp.48-49.

松崎寛・今井新悟・古川雅子「音声認識技術を活用した日本語韻律学習システムの開発と改良(オーガナイズドセッション「音声

情報処理技術を利用した第二言語の教育・学習支援」,合成,生成,韻律,音声一般)』『電子情報通信学会技術研究報告』査読無, SP, 音声113(308), 2013, pp.59-63

〔学会発表〕(計 2件)

古川雅子・Son Pham Thanh・今井新悟(2013)「日本語学習者のための音声認識技術を用いた発話練習教材の開発」教育システム情報学会, 2013年1月12日, 東北大学

松崎寛(2012)「音声認識システムを活用した発音自学システムの開発と試用」2012年日本語教育国際研究大会(ICJLE2012), 2012年8月18日, 名古屋大学

〔図書〕(計 2件)

松崎寛「イントネーション/音調/強調/促音/長音/拍(モーラ)/撥音/プロミネンス/ポーズ」(分担執筆)佐藤武義・前田富祺他編『日本語大事典』朝倉書店, 2014, Pp.2299.

畑佐一味・畑佐由紀子・百濟正和・清水崇文編『第二言語習得研究と言語教育』くろしお出版, 2012, Pp.309.

〔その他〕

ホームページ等

<http://www.u.tsukuba.ac.jp/~matsuzaki.hiroshi.fp/>

6. 研究組織

(1)研究代表者

松崎 寛 (MATSUZAKI, Hiroshi)

筑波大学・人文社会系・准教授

研究者番号: 10250648

(2)研究分担者

高橋 恵利子 (TAKAHASHI, Eriko)

広島大学・大学院教育学研究科・助教

研究者番号: 30710868

畑佐 由紀子 (HATASA, Yukiko)

広島大学・大学院教育学研究科・教授

研究者番号: 40457271

今井 新悟 (IMAI, Shingo)

筑波大学・人文社会系・教授

研究者番号: 50346582

古川 雅子 (FURUKAWA, Masako)

筑波大学・留学生センター・研究員

研究者番号: 20617287