

話者適応と文法誤りモデリングを用いた外国語教育システムの開発

著者	伊藤 彰則
URL	http://hdl.handle.net/10097/39578



話者適応と文法誤りモデリングを用いた 外国語教育システムの開発

(課題番号 16300260)

平成16年度～平成19年度科学研究費補助金
(基盤研究(B))

研究成果報告書

平成20年5月

研究代表者 伊藤 彰則

(東北大学 大学院工学研究科 准教授)

話者適応と文法誤りモデリングを用いた
外国語教育システムの開発
(課題番号 16300260)

平成16年度～平成19年度科学研究費補助金（基盤研究(B)）

研究成果報告書

平成20年5月

研究代表者 伊藤 彰則

(東北大学 大学院工学研究科 准教授)

はしがき

外国語の学習は根気の要る作業であるが、その外国語を操る能力の必要性はますます高まっている。日本語母語話者にとっては英語を初めとする外国語を学ぶことが必要であり、それ以外の母語話者にとっては日本語の学習が必要となっている。例えば、代表的な英語の能力試験の一つである TOEIC の受験者数は平成 15 年度には約 150 万人に達しており、勉強・仕事だけでなく日常生活にも外国語能力は必要不可欠になってきている。

外国語を操る能力は、主に「読む」「書く」「聞く」「話す」の四つからなるといわれる。これらのうち、「読む」「書く」については従来の教室で行う外国語教育でカバーできる範囲であり、基本的には文法学習に加えて「たくさん読んで、たくさん書く」という学習法がある程度確立している。「聞く」については、近年その重要性が認識され、日本においては中学・高校の学校教育やセンター試験などの入学試験にも導入されている。

「聞く」学習については、基本的に「たくさん聞く」という練習が主体となる。古くはカセットテープ、現在では CD からマルチメディア教材へとその主役が移り変わってきてはいるが、「流れてくる音声を聞き取る」という練習が主体であることは以前から大きく変わってはいない。

これらに対して、「話す」能力の教育は著しく遅れているといわざるを得ない。「話す」能力は、その他の三つの能力にも増して「体で覚える」ための訓練が必要であるのに、学校教育においては教室全員で行う発音訓練や簡単なロールプレイ以上のことは行われていない。「話す」能力の難しさは、正解との照合が難しい点にある。「読む」「書く」「聞く」については、正解がほぼ一意に決まる問題が作成でき、学習者はその問題を解く形で学習を進めていく。ところが、「話す」ための問題を作成したとしても、話し方が正しいのかどうかは母語話者しか基本的には判定できない。学校教育においては ALT など外国語母語話者を補助教師として使う試みがなされてはいるが、「たくさん話す」というレベルには程遠い。この問題を本質的に解決するためには、現状では英会話教室などに通って学習するしかなく、時間的・経済的に負担が重い。

音声認識技術を使う外国語学習システム(Computer Assisted Language Learning, CALL)は、この問題を根本的に解決できる可能性がある。手軽に使える外国語学習ソフトが多くの人に望まれていることは、任天堂 DS のソフト「えいご漬け」が 200 万本近く売れている

(平成 19 年 4 月現在) ことからわかる。音声認識を CALL システムに応用することによって、単語の発音練習、文のイントネーションの練習、会話練習など、これまで一人で行うことが難しかった「話す」学習の可能性が開けるのである。

音声 CALL システムを扱った研究は数多いが、いくつかの点で実用化にはまだ遠いと思われる点があった。一つは、発音評価の性能である。現在の音声認識は統計的なモデル化をその基礎にしており、多くの音声データを用いてモデルを学習する。この方法は、標準的な声質の話者に対しては的確な評価を可能にするが、声質が標準から外れている場合には的確な評価ができない。すなわち、「発音が悪い」と「声質が標準から外れている」ことの区別をすることが難しい。二つ目は、イントネーションを的確に評価する手法が確立していないことである。イントネーションの評価は発音の評価よりも難しく、また母語話者であっても人による揺らぎが大きいので、ある学習者のイントネーションがどの程度良いのかを機械的に判定することは難しい課題である。三つ目の問題点は、「話す」練習の最終目標である会話の練習ができない点である。「話す」というプロセスは、話したい内容から外国語の表現をリアルタイムに想起して、それを母国語とは異なるアクセント・

発音を使つて的確に発声するという一連の動作であつて、頭による学習というよりスポーツに近い。そのため、作文・単語の発音・イントネーションといった個々の練習をするだけでは不十分で、一連の動作を全てつないだ実際の発話の練習を大量にこなすことが必要不可欠である。しかし、従来の CALL システムは、会話に近いことができるシステムであつても、いくつかの決められた文の一つを読み上げる形式であつたり、あるいはきわめて限られた表現しか発話が受け付けられないといった、限定された形の学習しか許されていなかった。

本報告は、これらの点を解決し、本当に使える音声 CALL システムを実現すべく、平成 16 年度から 19 年度の 4 年間にわたつて取り組んだ研究成果をまとめたものである。先にあげた三つの問題点に対応して、「発音評価の高精度化」「イントネーションの評価」「会話型 CALL システム」の三つの課題に取り組み、完全とはいえないまでも一定の成果を収めることができたと考えている。今回の研究では、実際にシステムを構築してユーザー評価を行うまでには至らなかったが、今後本研究で開発した技術を実際の環境で評価する必要があるであろう。

平成 20 年 3 月

研究代表者 伊藤彰則

研究組織

研究代表者 伊藤 彰則 (東北大学 大学院工学研究科 准教授)
研究分担者 牧野 正三 (東北大学 大学院工学研究科 教授)
研究分担者 鈴木 基之 (東北大学 大学院工学研究科 助教)

交付決定額 (配分額)

金額単位：円

交付年度	直接経費	間接経費	合計
平成16年	3,200,000	0	3,200,000
平成17年	1,400,000	0	1,400,000
平成18年	1,400,000	0	1,400,000
平成19年	1,400,000	420,000	1,820,000
計	7,400,000	420,000	7,820,000

研究発表

1. 雑誌論文

- (1) Oh Pyo Kweon, Motoyuki Suzuki, Akinori Ito and Shozo Makino, "An evaluation method of Japanese pronunciation for Korean native speakers," Educational Technology Research, 27, 1-2, 9-16, 2004. (査読有)
- (2) Oh-Pyo Kweon, Akinori Ito, Motoyuki Suzuki and Shozo Makino, "A Grammatical Error Detection Method for Dialogue-based CALL system," Journal of Natural Language Processing, 12, 4, 137-156, 2005. (査読有)
- (3) Akinori Ito, Tadao Nagasawa, Hirokazu Ogasawara, Motoyuki Suzuki and Shozo Makino, "Automatic detection of English mispronunciation using speaker adaptation and automatic assessment of English Intonation and rhythm," Educational Technology Research, 29, 13 - 23, 2006. (査読有)
- (4) Akinori Ito, Yen-Ling Lim, Motoyuki Suzuki and Shozo Makino, "Pronunciation error detection for computer-assisted language learning system based on error rule clustering using a decision tree," Acoustical Science and Technology, 28, 2, 131-133, 2007. (査読有)
- (5) 伊藤彰則, 今野樹, 鈴木基之, 牧野正三: 「決定木を用いた単語クラスタリングによる英語韻律自動評価の高精度化」, 電子情報通信学会論文誌(D), J91-D, 2, 358-366, 2008. (査読有)
- (6) Oh-Pyo Kweon, Akinori Ito, Motoyuki Suzuki and Shozo Makino, "A dialogue-based CALL system for Japanese conversation," Proceedings of the 18th International Congress on Acoustics, vol. III, pp. 2015-2018, 2004. (査読有)
- (7) Oh Pyo Kweon, Akinori Ito, Motoyuki Suzuki and Shozo Makino, "A Japanese dialogue-based CALL system with mispronunciation and grammar error detection," Proceedings of International Conference on Spoken Language Processing, vol. III, pp. 1833-1836, 2004. (査読有)
- (8) Motoyuki Suzuki, Hirokazu Ogasawara, Akinori Ito, Yuichi Ohkawa and Shozo Makino, "Speaker Adaptation Method for CALL Systems Using Bilingual Speakers' Utterances," Proceedings of International Conference on Spoken Language Processing, vol. IV, pp. 2929-2932, 2004. (査読有)
- (9) Oh Pyo Kweon, Akinori Ito, Motoyuki Suzuki and Shozo Makino, "A Japanese dialogue-based CALL system to detect grammar errors of learners," Proceedings of the 2nd Student-Organizing International Mini-Conference on Information Electronic System, pp. 293-296, 2004. (査読無)
- (10) Akinori Ito, Yenling Lim, Motoyuki Suzuki and Shozo Makino, "Pronunciation Error

Detection Method Based on Error Rule Clustering Using a Decision Tree,” Proceedings of 9th European Conference on Speech Communication and Technology, pp. 173-176, 2005. (査読有)

- (11) Ryohei Tsutsui, Motoyuki Suzuki, Akinori Ito and Shozo Makino, “Language Modeling towards Speech Recognition for Voice-interactive CALL System,” Proceedings of 2nd International Workshop of Tohoku University and Yeungnam University, pp. 29, 2005. (査読無)
- (12) Motoyuki Suzuki, Tatsuki Konno, Akinori Ito, Shozo Makino, “Automatic Evaluation System of English Prosody for Japanese Learner's Speech,” Proceedings of 5th International Conference on Education and Information Systems, Technologies and Applications (EISTA), 2007. (査読有)
- (13) Tomoaki Konno, Tatsuki Konno, Motoyuki Suzuki, Akinori Ito and Shozo Makino, “Characteristic Analysis of a Prosodic Evaluation Tendency by English Native Speakers,” Proceedings of 3rd International Workshop of Tohoku University and Yeungnam University, pp. 41-42, 2006. (査読無)

2. 学会発表

- (1) 権五杓, 鈴木基之, 伊藤彰則, 牧野正三: 「対話型 CALL システムのための文法誤り検出法」, 日本音響学会, 2005年3月17日, 東京農工大学 (小金井市)
- (2) 鈴木基之, 林燕玲, 伊藤彰則, 牧野正三: 「発音誤り検出のための誤り規則のクラスタリング」, 日本音響学会, 2005年9月29日, 東北大学 (仙台市)
- (3) 筒井良平, 鈴木基之, 伊藤彰則, 牧野正三: 「日本人英語音声認識のための音響モデルに関する検討」, 電気関係学会東北支部連合大会, 2006年8月31日, 岩手大学 (盛岡市)
- (4) 今野樹, 鈴木基之, 伊藤彰則, 牧野正三: 「単語の重要度に基づく英語母語話者による韻律評価傾向の特徴分析」, 日本音響学会, 2006年9月13日, 金沢大学 (金沢市)
- (5) 筒井良平, 鈴木基之, 伊藤彰則, 牧野正三: 「英語 CALL システムのための文法誤りを含む発話のモデル化」, 日本音響学会, 2007年3月13日, 芝浦工業大学 (江東区)
- (6) 今野樹, 鈴木基之, 伊藤彰則, 牧野正三: 「英語韻律評価のための言語情報による重要単語クラスタリング」, 日本音響学会, 2007年3月14日, 芝浦工業大学 (江東区)
- (7) 今野智明, 鈴木基之, 伊藤彰則, 牧野正三: 「合成音声を用いた英語のイントネー

ション評価の実用性に関する検討」，電気関係学会東北支部連合大会，2007年8月23日，弘前大学（弘前市）

- (8) 筒井良平，鈴木基之，伊藤彰則，牧野正三：「生成テキストの Ngram を用いた英語学習者発話の音声認識」，日本音響学会，2007年9月19日，山梨大学（甲府市）
- (9) 筒井良平，鈴木基之，伊藤彰則，牧野正三：「生成したテキストの Ngram を用いた英語学習者による文法誤りを含んだ発話の音声認識の高精度化」，音声言語シンポジウム，2007年12月20日，NTT 京阪奈（京都府）

研究成果の概要

1. 発音評価の高精度化

発音評価は、単語や文を単位とした発声における、分節的な良さを評価するものである。日本人のための英語学習システムであれば、例えば/l/と/r/や/ah/と/ae/の区別の的確さを測ることになる。評価方法としては、全体としての良さを数値として表現する方法（例えば「あなたの発音は100点満点中70点です」等）、あるいは発音の誤りを検出して指摘する方法（例えば「“apple”の/a/の発音が間違っています」等）が考えられる。本研究では、後者の方法を主に採用する。

発音の誤り検出の基本方式として、河合ら(2001)による二言語音響モデルを用いた手法を用いる。この方法は、ある単語や文について、正しい対象言語の発音を表す音響モデルと、それに対して誤って発声されるであろう学習者母語の音響モデルを並列に連結し、実際の発声がどちらに近いのかを測る手法である。このように二言語の音響モデルを並列に連結した単語／文モデルを用いることには、二つの問題点がある。一つは、音響モデルによって発音を評価する場合、「発音の近さ」と「話者の近さ」を区別することができないという点である。ある学習者の声質が、二言語の音響モデルを学習したデータベースのどちらからも遠かった場合、それらの音響モデルによる評価値は信頼性がなく、評価結果の精度は低下する。もう一つの問題は、「どの程度対象言語から遠ければ誤りと判定するか」という基準が文脈に依存することである。ある文脈では発音が少し悪くなっただけでも全体の「発音の良さ」の印象が低下する一方、別な文脈ではかなり悪い発音であっても全体の印象にはあまり影響しない、といったことが起こりうる。

前者の問題に対しては、音声認識で広く研究されている話者適応の技術を応用することで解決する。しかし、例えば英語を学習する日本語母語話者の例で考えた場合、そのような学習者は、日本語を正しく発声することはできるが、英語を正しく発声することはそもそも期待できない。そのため、従来の話者適応法では、日本語の音響モデルは適応できるが、英語の音響モデルは適応できない。そこで、本研究では、ある一つの言語による適応データのみを使って、二つの言語の音響モデルを同時に適応する話者適応手法を開発した。まず、発声データが不完全であっても適応が可能なVFS(Vector Field Smoothing)法を用いて、同じ音響空間内に布置された日本語と英語の音素を同時に適応することによって、日本語だけでなく英語の音響モデルの適応も可能であることを明らかにした。次に、同じく不完全な発声データに対して適応可能なMLLR(Maximum Likelihood Linear Regression)法を用いて適応を行い、性能の向上を得ることができた。次に、SAT(Speaker Adaptive Training)法を用いて、最初から話者適応に適した二言語の音響モデルを作成しておくことによって、通常作成した音響モデルをVFS法やMLLR法で適応するよりもより高い適応効果が得られることを明らかにした。

後者の問題に対しては、音素環境や文の種類、音素の文内での位置などを決定木によっ

てクラスタリングし、クラスタごとに最適な誤り検出基準を設定することによって性能を向上させる手法を開発した。この方法では、あらかじめ用意された質問（「前の音素は/r/か？」等）に対して最適な閾値でクラスタを分割し、最も誤り検出性能が高かった質問で再帰的に文脈をクラスタリングしていく。最終的に、最適な文脈の分類とそのときの最適な閾値を同時に求めることができる。最終的に、90%以上の誤り検出精度を得ることができた。

2. イントネーション・リズムの評価

イントネーションとリズムは外国語学習において重要な要素であるが、自動評価の研究は少なく、また精度も限られたものであった。我々はこの問題に取り組み、どのような特徴量が評価に有効かという基礎的な検討から、イントネーション・リズム評価手法まで一貫して研究・開発を行った。

イントネーションは主に意味的・文法的まとまりである韻律句全体を流れるピッチの流れとして定義される。英語の場合、主に音声の基本周波数の変化と捉えることができるが、基本周波数だけでなく音声の強さ（音声パワー）も影響する。また、基本周波数の高さには個人差があるので、これを正規化する手法が必要になる。

まずイントネーション評価の基本的な手法として、単語ごとにイントネーション特徴量を算出し、それを教師話者の発声と比較し、単語ごとの特徴量の距離を文全体にわたって積算することで分のイントネーション評価値を計算する手法を開発した。ここで我々は、イントネーション評価の特徴量について検討し、対数基本周波数およびその時間微分、対数パワーおよびその時間微分を特徴量として用いるのがよいということを見出した。次に、イントネーション評価の厳密さが単語によって異なることに我々は注目した。そこで、単語ごとに評価の厳密さを制御することによって文全体の評価値の精度を上げる手法を開発した。ここで、ある単語についてどのように評価の厳密さを与えるかが大きな問題となる。そこで我々は、単語をその品詞と文脈によってクラスタリングし、クラスタに対して最小二乗法によって最適な重みを与える手法を開発した。クラスタリングの手法として、決定木を用いる。決定木によるクラスタリングを使うことで、実際の利用時には学習データに出現しなかった単語に対しても頑健なクラスタの決定が可能になる。日本語母語話者190名が発話した英語発音に対して、提案法でイントネーション評価を行ったところ、英語母語話者の評定値との相関が最大0.45となった。複数の評定者間の評定値の相関が0.5~0.6なので、これに迫る性能を得ることができた。

リズム評価については、従来ほとんど行われていなかった研究であるので、まず特徴量について検討した。特徴量として、学習者の単語の持続時間と教師の単語の持続時間の比（単語持続時間比）、学習者と教師の単語音声をDPマッチングによって対応付けたときの距離（DP距離）、単語持続時間比の分散、1文の中でのある単語の持続時間の割合（単語相対持続時間）、韻律句以外の箇所は無音が挿入されている比率（無音挿入誤り率）等について検討した。その結果、単語持続時間比が最も良い特徴量であることが確かめら

れた。日本語母語話者 190 名が発話した英語発音について、提案法でリズム評価を行ったところ、英語母語話者の評定値との相関は最大 0.55 であった。評定者間の評定値の相関が 0.58~0.65 であるので、これに近い精度を得ることができた。

3. 対話型 CALL システムの開発

コンピュータと学習者が会話をしながら発話の練習をするシステムは、音声 CALL システムの究極の姿である。しかし、対話には音声認識を使う以上、認識誤りを避けることができない。また、音声認識システムは、ユーザの発話が音声学的・文法的に正しいということの大前提として構築されており、外国語学習者のように発音にも文法にも誤りを含んだ発話を、学習者の意図通りに認識することはきわめて難しい。そのため、対話型 CALL システムを構築するためには、誤りを含んだ発話をいかにして（発話者の犯した誤りを含めてそのまま）認識するかが最大の課題となる。

これに対して、一般的に「どのような発話をしても良い」という状況で音声認識を高精度で行うことはほとんど不可能である。一方、学習者の発話を制限しすぎるとは、対話型 CALL システムの意義が失われる。そこで、対話をする前に、特定の語彙や文法項目について学習者にまず事前学習をしてもらい、システムとの対話ではそれに関連する（しかし、まったく同じではない）内容について対話を行ってもらおうという方法を提案した。日本語を学習する韓国語母語話者に対してこの方法を適用し、事前学習がない場合と比較した結果、事前学習を行うと、システム側が用意した認識文法から大きく外れる発話（文法外発話）の割合が大きく減少することがわかった。

次に、このように状況を制限することでシステム側の意図に沿った学習者発話を得られたとしても、学習者の発話には文法的な誤りが含まれる。そこで、文法的な誤りを含めて認識を行うことが必要となる。従来の対話型 CALL システムでは、どのような誤りがあるかを事前に専門家が予測し、そのような誤りを含めて発話を受理する認識文法をあらかじめ用意していた。しかし、この方法では、新たな場面に対して文法を用意するための手間が膨大であり、実用的ではない。そこで、一般的に学習者が犯すであろう誤りを「誤りルール」という形で事前に用意し、ある対話の場面で受理できる（誤りを含まない）表現に対して誤りルールを適用することで文法を拡大する手法を提案した。この手法を、韓国語母語話者による日本語発声に適用し、有効性を確認した。

この方法では、誤りルールは事前に用意しておく必要がある。そこで我々は、コーパスから誤りルールを自動的に抽出する手法を検討した。また、音声認識の精度を上げるため、有限状態オートマトンではなく N-gram を使って認識を行う手法を検討した。N-gram を用いることで認識性能は向上したが、認識精度はまだ十分ではなかった。そこで、学習者の発話が「その場面で対象言語母語話者が発生するであろう正しい表現」（正解文）からそれほど大きくは外れていないという性質に注目し、正解文との距離を参照してより正しい（学習者の発話に近い）文を認識する手法を開発した。日本人の発生した英語文に対して

この手法を適用し、80%以上の単語認識率を得た。

目次

1. 発音評価の高精度化

- An evaluation method of Japanese pronunciation for Korean native speakers.....1
- Automatic detection of English mispronunciation using speaker adaptation and automatic assessment of English Intonation and rhythm.....9
- Pronunciation Error Detection Method Based on Error Rule Clustering Using a Decision Tree.....20
- Speaker Adaptation Method for CALL Systems Using Bilingual Speakers' Utterances...24
- 発音誤り検出のための誤り規則のクラスタリング.....28
- Pronunciation error detection for computer-assisted language learning system based on error rule clustering using a decision tree.....30

2. イントネーション・リズムの評価

- Characteristic Analysis of a Prosodic Evaluation Tendency by English Native Speakers33
- Automatic Evaluation System of English Prosody for Japanese Learner's Speech.....34
- 単語の重要度に基づく英語母語話者による韻律評価傾向の特徴分析.....40
- 英語韻律評価のための言語情報による重要単語クラスタリング.....42
- 決定木を用いた単語クラスタリングによる英語韻律自動評価の高精度化.....44
- 合成音声を用いた英語のイントネーション評価の実用性に関する検討.....53

3. 対話型 CALL システムの開発

- A dialogue-based CALL system for Japanese conversation55
- 対話型 CALL システムのための文法誤り検出法.....59
- A Japanese dialogue-based CALL system with mispronunciation and grammar error detection.....61
- A Japanese dialogue-based CALL system to detect grammar errors of learners.....65
- A Grammatical Error Detection Method for Dialogue-based CALL system.....69
- 日本人英語音声認識のための音響モデルに関する検討.....89
- Language Modeling towards Speech Recognition for Voice-interactive CALL System.....90
- 英語 CALL システムのための文法誤りを含む発話のモデル化.....91

- 生成テキストの Ngram を用いた英語学習者発話の音声認識.....93
- 生成したテキストの Ngram を用いた英語学習者による文法誤りを含んだ発話の音声認識の高精度化.....95

1. 発音評価の高精度化

本報告書収録の学術雑誌等発表論文は本ファイルに登録していません。なお、このうち東北大学在籍の研究者の論文で、かつ、出版社等から著作権の許諾が得られた論文は、個別に **TOUR** に登録しております。