社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

識別モデルを用いた英語文発声からの強勢自動検出

加藤　集平†　鈴木　雅之††　峯松　信明††　広瀬　啓吉†　山内　豊†††
西川　惠††††

† 東京大学大学院情報理工学系研究科 〒113–0033 東京都文京区本郷 7–3–1
†† 東京大学大学院工学系研究科 〒113–8656 東京都文京区本郷 7–3–1
††† 東京国際大学商学部 〒350–1197 埼玉県川越市的場北 1–13–1
†††† 東海大学外国語教育センター 〒259–1292 神奈川県平塚市北金目 1117
E-mail: †,††{kato,suzuki,mine,hirose}@gavo.t.u-tokyo.ac.jp, †††yyama@tiu.ac.jp

あらまし　学習者が英語の自然な発音を身につけるためには，/l/と/r/の区別に代表される分節的特徴以外にも，強勢やリズム，イントネーションといった韻律的特徴を修得する必要がある．本研究では強勢，特に文強勢に注目し，読み上げ音声からの文強勢検出を目指した．具体的には，評価対象音声に対して音節ごとに強勢の度合いを識別モデルを用いて自動推定を行った．評価対象音声には日本人学生による英語読み上げ音声データベースを用いたが，本データベースには音節ごとに強勢の度合いを判定したラベルは付与されておらず，手動判定によりラベルを付与した．また，読み上げ音声を使用するため，評価対象音声の発話内容は既知である．そこで，音響特徴量以外の，テキストから得られる特徴量も積極的に使って推定精度を上げることを目指した．実験の結果，弱勢，強勢，核強勢の 3 段階の検出において，最高で 81.8%の正解率が得られたほか，テキストから得られる特徴量も有効であることが分かった．
キーワード　識別モデル，強勢，韻律，発音評価，外国語学習支援

Automatic detection of English sentence stress using discriminative models

Shuhei KATO†, Masayuki SUZUKI††, Nobuaki MINEMATSU††, Keikichi HIROSE†, Yutaka YAMAUCHI†††, and Megumi NISHIKAWA††††

† Graduate School of Information Science and Technology, The University of Tokyo
7–3–1, Hongo, Bunkyo, Tokyo, 113–0033, Japan
†† Graduate School of Engineering, The University of Tokyo
7–3–1, Hongo, Bunkyo, Tokyo, 113–8656, Japan
††† Department of Business and Commerce, Tokyo International University
1–13–1, Matobakita, Kawagoe, Saitama, 350–1197, Japan
†††† Foreign Language Center, Tokai University 1117, Kitakaname, Hiratsuka, Kanagawa, 259–1292, Japan
E-mail: †,††{kato,suzuki,mine,hirose}@gavo.t.u-tokyo.ac.jp, †††yyama@tiu.ac.jp

Abstract　To learn natural English pronunciation, in addition to learning segmental features acquisition of prosodic features such as stress, rhythm, and intonation is very important. In this paper, we focused on recognizing sentence stress level for use in a computer assisted language learning application. Specifically, we used a discriminative model to estimate the stress level for each syllable in the target speech. The target speech was selected from the English Read by Japanese (ERJ) database. As the ERJ database does not supply stress level labels, we had them labeled manually. Since read speech was used, we made use of grammar-based features in order to improve the accuracy of estimation. In the case of 3-levels of stress (no stress, stress, stress nucleus), the recognition accuracy was 81.8%, and the grammar-based features improved the accuracy.
Key words　discriminative model, stress, prosody, automatic evaluation, CALL

1. はじめに

国際化が進む中で，英語による音声コミュニケーションを行う機会は増加している．最近では，社内公用語を英語とする日本企業も現れてきた．そのような状況で，語学教育も盛んになってきている．例えば，2011 年度から小学 5，6 年生での外国語活動が必修となった．この外国語活動では，外国語を用いてコミュニケーションを図る楽しさを体験する，積極的に外国語を聞いたり，話したりすることを指導することとされ，話し言葉としての外国語教育が重視されている [1]．また，多くの成人向け外国語教室でも，話し言葉によるコミュニケーションを重視した教育が行われている．このように，近年の語学教育においては，聞く・話すことによるコミュニケーション能力の向上を目標とした教育が重視されている．しかし，そのようなコミュニケーションを中心とした語学教育を十分に行える能力を持った教師は満足に確保されているとは限らない．例えば，小学校の外国語活動の現状について日本英語検定協会が行った調査によると，外国語活動における問題・課題であると感じていることとして,「指導者（担当教員）の質・技術」が，教育委員会を対象とした調査では 15 項目中 1 番目に，公立小学校を対象とした調査では 16 項目中 3 番目にあげられている [2, 3]．十分な教育能力を持った教師が不足している状況では，いくらコミュニケーション能力の向上を教育目標としたところで，実効性は限られたものになってしまうであろう．

そこで，教師不足をコンピュータによって補う，Computer Assisted/Aided Language Learning (CALL) システムが数多く開発されている [4–8]．学習者の音声から音響特徴量を抽出し，あらかじめ用意しておいたモデルと比較して，学習者にフィードバックするという形態が一般的である．

既存の CALL では，/l/や/r/などの音素が正しく発音できているかといった，分節的特徴の評価を目的としたものが多く，強勢やリズム，イントネーションといった韻律的特徴の評価を目的としたものは少ない．そこで，本研究では強勢，特に文強勢に注目し，読み上げ音声からの自動検出を行うことを目指す．文強勢は自然な英語リズムを形成するほか，強調などの話者の意図が表現されるため，母語話者と十分な意思疎通を行う上で重要な要素だからである [9]．文強勢検出を目的とした研究は既に存在する [10–13]．評価対象音声も研究によって様々であるが，例えば [11] は,「日本人学生による読み上げ英語音声データベース」（ERJ (English Read by Japanese) データベース）[14] の「文強勢，文リズムに関する文」（以下リズム文とする）に対する読み上げ音声を用いて文強勢検出を行っている．この音声の特徴は，英語の強勢・リズムに着目して構成された文を，原稿に記載された強勢記号に従って学習者自身が読み上げたもので，学習者自身は正しく読めたと考えている音声となっている点である．読み上げ文が英語の強勢・リズムに着目して作られている点，学習者自身が正しく読めたと考えている音声である点で教育的な観点から評価対象としてふさわしく，読み上げ音声という点で技術的にも扱いやすい．ただし, [11] では HMM を使って評価しているため，特徴量を増やしたりすることに限界がある．そこで，本研究では ERJ のリズム文の読み上げ音声を評価対象として，識別モデルを用いてより精度の高い文強勢の自動評価を目指すことにする．具体的には，評価対象音声に対して音節ごとに強勢の度合いを高精度に自動推定することを目的とする．ただし，ERJ には音節ごとに強勢の度合いを判定したラベルは付与されていないため，手動判定という形でラベルを付与した．また，本研究では読み上げ音声を使用するため，評価対象音声の発話内容は既知である．そこで，音響特徴量以外の，テキストから得られる特徴量も積極的に使って推定精度を上げることを目指した．

2. 音声試料

評価対象の音声試料は，ERJ のリズム文である．ERJ は，日本語を母語とする大学生・大学院生相当の学生による読み上げ音声データベースである．リズム文の読み上げ原稿には，音節ごとに強勢の度合いを示す記号が 3 段階（弱勢，強勢，核強勢）で付けられている．ERJ の音声収録にあたっては，発声者は事前の発声練習が許されており，収録も発声者自身が「正しい」と判断できる発声が得られるまで繰り返されている．

ERJ のリズム文は全部で 120 文あるが，本研究では，そのうち第五著者（英語教師）が「発声者によって出来に差がつきやすい」と判断し選んだ 20 文を使用した．この 20 文について，すべての発声を評価対象とした．結果，評価対象となったのは 684 発声（7,414 音節），話者 202 名（男性 100 名，女性 102 名）である．

以下の実験においては，全 7,414 音節を話者および発声対象文が重複しないようにほぼ 3 等分に分割し，2 つを学習データ，残りの 1 つをテストデータとした．

3. 手動判定による強勢ラベリング

3.1 評価者

評価者は 2 名である．英語音声学を専攻している大学院博士課程の学生（以下評価者 A）と第六著者（英語教師，以下評価者 B）である．両者で事前に判定基準の統一をとることなく，それぞれ独立に判定を行った．

3.2 手動判定タスク

実験で用いる発声を聴取して，音節ごとに強勢の度合いを判定した．音声収録時の強勢記号は 3 段階（弱勢，強勢，核強勢）であるので，各々に対して，より弱い／適切／より強い，の 3 段階を考慮し，合計 9 段階とし（1 が最も弱く，9 が最も強い），また,「発声されていない」ことを示す 0 を加えた．以下，9 段階を 3 段階に丸めて用いる場合（1 から 3 は「弱勢」としてひとくくりにする．以下同様）を「3 段階」，9 段階をそのまま用いる場合を「9 段階」と呼ぶことにする．

全 7,414 音節に対する手動判定の結果を表 1 および表 2 に示す．参考のため，収録時に読み上げ原稿に呈示された強勢記号の個数を表 1 に併記しておく．

また，手動判定の安定性を検証するために，評価対象である 684 発声のうち 64 発声を選び，数週間の時間を置いて再判定を行った．再判定対象発声の選定は，1 回目の判定順（684 回

表 1　手動判定の結果（3 段階）

評価	0（発声されていない）	1–3（弱勢）	4–6（強勢）	7–9（核強勢）
評価者 A	26	4,861	3,101	506
評価者 B	31	5,022	2,158	1,283
（原稿の強勢記号）	0	5,504	1,395	1,595

表 2　手動判定の結果（9 段階）

評価	0	1	2	3	4	5	6	7	8	9
評価者 A	19	86	2,438	1,639	1,262	1,189	347	181	243	10
評価者 B	24	80	3,847	358	488	654	858	771	323	11

表 3　評価者内の判定の一致率および相関

評価者	段階	一致率	相関
評価者 A	3 段階	0.865	0.793
評価者 A	9 段階	0.618	0.869
評価者 B	3 段階	0.860	0.868
評価者 B	9 段階	0.651	0.911

表 4　評価者間の判定の一致率および相関

段階	一致率	相関
3 段階	0.775	0.776
9 段階	0.436	0.829

の判定作業で何番目に判定を行ったか）のバランスを考慮して，準ランダムに行った．判定条件は 1 回目と同様であるが，1 回目に自分が下した判定は参照させていない．1 回目の判定と再半知恵の一致率および相関を表 3 に示す．

両評価者は事前に判定基準の打ち合わせ無しに判定を行ったが，評価者間の（1 回目の）判定の一致率および相関を表 4 に示す．SVM を用いた自動判定を試みるが，その精度評価は，表 3 および表 4 が比較対象となる．

4. 先行研究の手法を用いた実験

1. で，先行研究として HMM を用いた手法を紹介した [11–13]．本研究ではベースラインとして，[11] と同様の手法を実装し，実験を行った．なお，本節での実験は 3 段階の場合についてのみ行った．

[11] では，音節の強勢／弱勢を HMM でモデル化し，入力文発声の各音節に対して HMM を用いて強勢／弱勢の判定を行っている．本研究でもこれにならい，弱勢／強勢／核強勢の 3 クラスのモデルを構築した（手動判定が 0 であったものについては，本実験では使用しなかった）．特徴量は SPTK を用いて抽出した [15]．音響分析条件を表 5，HMM の学習条件を表 6 に示す．

特徴量も [11] にならった．F_0 は，SPTK によって抽出した F_0 を母音区間だけ残して対数をとったものを線形補間する形で求めた．文頭（末）の無声区間の F_0 は，母音区間の F_0 の対数をとったものから平均 F_0 と標準偏差 σ を求め，母音区間から平均音節時間長前（後）に，平均 $F_0 - 3\sigma$ の値（基本周波数生成過程モデルにおける基底周波数 F_b に相当）にあるものとして線形補間を行った．さらに，F_0 とエネルギーに対しては，当該発声の平均値を引く形で正規化を行ったものを最終的な特徴量として用いた．

表 5　音響分析条件

標本化周波数	16 kHz
量子化ビット数	16 bit
抽出周期	8.0 ms (128 samples)

表 6　HMM の学習条件

特徴量	対数 F_0 + Δ対数 F_0, エネルギー+Δ エネルギー, MFCC+ΔMFCC（各 1〜4 次元）
トポロジー	6 状態 4 分布，left-to-right
分散行列	3 種類の特徴量に対して個別に全共分散行列を算出

表 7　先行研究の手法を用いた実験の結果

学習データ	テストデータ	正解率
評価者 A	評価者 A	0.444
評価者 B	評価者 B	0.426

5. 先行研究の手法を用いた実験の結果

実験結果を表 7 に示す．評価者 A，評価者 B いずれのラベルを用いた場合でも，正解率はほぼ等しくなった．表 1 に示すように，強勢，弱勢の出現には数的偏りがある．先行研究ではこれらの事前分布を使わずに評価を行っており，表 7 の結果は，常に「弱勢である」と答えた場合よりも，正解率が劣る結果となった．

6. 提案手法を用いた実験

提案手法では，SVM を用いて強勢の度合いを判定する．具体的には，R の kernlab パッケージに含まれる ksvm を用い

表 8　提案手法で用いた特徴量セット

特徴量セット	特徴量
a（音響特徴量）	母音部平均対数 F_0 ／母音部継続長／母音部平均エネルギー／母音部平均 MFCC（12 次元）（いずれも発話ごとに正規化），母音部平均対数 F_0 ／平均エネルギーの差（前後 2 音節まで），母音部継続長の比（前後 2 音節まで）
b（コンテキスト）	音節中の母音が単母音か否か，フレーズ頭の単語（最初の 1 単語）／フレーズ中の単語／フレーズ末の単語（最後の 1 単語）中の音節である，内容語中／機能語中の音節である，どの品詞中の音節か（名詞／代名詞／動詞／形容詞／副詞／前置詞／接続詞／間投詞），単語を単独発声したときに第 1 強勢／第 2 強勢／弱勢で発声される
c（原稿の強勢記号）	収録時に参照した強勢記号（弱勢／強勢／核強勢）
a-b	（a と b を合わせたもの）
a-c	（a と c を合わせたもの）
b-c	（b と c を合わせたもの）
a-b-c	（a, b, c すべてを合わせたもの）

た [16]．カーネルはガウスカーネルを使用し，ハイパーパラメータについては ksvm のデフォルトに設定した．手動判定が 0 のデータはここでも使用しなかった．使用した特徴量セットを表 8 に示す．

特徴量セット a に含まれるのは音響特徴量で，音節の母音部の平均 F_0，継続長，平均エネルギー，平均 MFCC（12 次元），およびそれらの差または比を前後 2 音節に対して計算している．このような特徴量を用いたのは，1) 強勢が，高さ，長さ，強さ，音質の複合的な変化により実現されることと，2) 強勢とは，他の音節よりも際立って聞こえることを指すからである [17, 18]．

特徴量セット b に含まれるのは，音節が持つコンテキストに関するシンボリックな特徴量である．これらは，どのような音節に文強勢が置かれやすいかという観点から選んでいる [17–19]．

特徴量セット c に含まれるのは，ERJ の原稿に記載されていた強勢記号に関する特徴量である．発声者は原稿に記載されていた強勢記号のとおりに発声できていると考えているため，有用な情報であることが期待される．

特徴量セット b, c には，音響的な情報が含まれておらず，ある読み上げテキスト中のある音節に対しては，どの発声者に対しても同じ特徴量となることに注意されたい．なお，特徴量セット b, c に関しては，各特徴量を 0/1 の二値で表現した（「X である」という特徴量は，X である場合は 1，X でない場合は 0 をとる）．最終的に，a, b, c のうち 2 つを組み合わせた特徴量セット，すべてを組み合わせた特徴量セットを用意した．

評価は，手動判定の評価者クローズドの場合と，評価者オープンの場合の両方を行った．すなわち，評価者クローズドの場合は評価対象の全 7,414 音節を発声者および発声対象文が重複しないように 3 分割したもののうち 2 つで学習，残り 1 つでテストするという作業を 3 回繰り返し平均をとった．評価者オープンの場合は一方の評価者のデータで学習し，他方の評価者のデータでテストを行った．

7.　提案手法を用いた実験の結果

提案手法を用いた実験の結果を表 9，表 10（3 段階）および表 11，表 12（9 段階）に示す．

まず，各特徴量セットに対する結果を比較する．3 段階の場合は，学習データとテストデータの組み合わせがいずれの場合でも，正解率および相関が最も低かったのは特徴量セット a であった．一方，最も高かった特徴量セットは，学習データとテストデータの組み合わせによって様々であった．4 つの組み合わせの結果を平均すると，正解率が最も高かったのは特徴量セット a-c（平均 0.786）であり，相関が最も高かったのは特徴量セット b-c および a-b-c（平均 0.737）であった．9 段階の場合は，学習データとテストデータの組み合わせがいずれの場合でも，正解率および相関が最も低かったのは特徴量セット a であり，3 段階の場合と同じ結果となった．一方，学習データとテストデータの組み合わせがいずれの場合でも，正解率が最も高かったのは特徴量セット a-c であった．相関については学習データとテストデータの組み合わせによって様々であったが，4 つの組み合わせの結果を平均すると，最も高かったのは特徴量セット a-c（平均 0.761）であった．以上をまとめると，特徴量セット a-c を用いた場合がおおむね最も高精度であったと言える．また，3 段階の場合の結果を表 7 に示した先行研究の結果と比べると，大幅に上回る精度が得られた．

特徴量セット a, b, c をそれぞれ単独で用いた場合の結果については，特徴量セット a のみを用いた場合が正解率，相関とも最も低かった．特徴量セット b と c の結果の差は小さかった．2 つの特徴量セットを組み合わせた場合 (a-b, a-c, b-c) の結果については，特徴量セット a, b, c をそれぞれ単独で用いた場合の結果と比べて正解率，相関ともおおむね向上していた．

次に，すべての特徴量セットの中でおおむね最も高精度であった特徴量セット a-c を用いた場合の結果と，手動判定の結果を比較する．評価者クローズドの結果については表 3 と比較することになる．3 段階に関しては，どちらの評価者のラベルを用いた場合でも，正解率，相関とも手動判定の一致率，相関をそれぞれ少し下回る結果となった．9 段階に関してもやはり，どちらの評価者のラベルを用いた場合でも手動判定の精度を下回った．しかし，評価者 B のほうが評価者 A よりも手動判定の精度に近い結果が得られた．評価者オープンの結果については表 4 と比較することになる．3 段階に関しては，正解率およ

表 9　提案手法による実験の結果（正解率）(3 段階)

学習データ	テストデータ	a	b	c	a-b	a-c	b-c	a-b-c
評価者 A	評価者 A	0.695	0.807	0.810	0.769	0.807	**0.818**	0.799
評価者 B	評価者 B	0.659	0.753	0.759	0.734	**0.796**	0.765	0.769
評価者 A	評価者 B	0.694	0.768	0.763	0.774	0.767	0.770	**0.776**
評価者 B	評価者 A	0.728	0.774	0.764	**0.785**	0.775	0.768	0.782

表 10　提案手法による実験の結果（相関）(3 段階)

学習データ	テストデータ	a	b	c	a-b	a-c	b-c	a-b-c
評価者 A	評価者 A	0.477	0.688	0.693	0.610	0.686	**0.712**	0.675
評価者 B	評価者 B	0.519	0.707	0.753	0.664	**0.773**	0.748	0.737
評価者 A	評価者 B	0.634	0.788	0.778	0.800	0.781	0.797	**0.805**
評価者 B	評価者 A	0.614	0.697	0.690	**0.733**	0.705	0.691	0.732

表 11　提案手法による実験の結果（正解率）(9 段階)

学習データ	テストデータ	a	b	c	a-b	a-c	b-c	a-b-c
評価者 A	評価者 A	0.398	0.431	0.462	0.434	**0.470**	0.445	0.455
評価者 B	評価者 B	0.546	0.595	0.618	0.592	**0.624**	0.610	0.614
評価者 A	評価者 B	0.432	0.478	0.475	0.459	**0.482**	0.478	0.468
評価者 B	評価者 A	0.349	0.366	0.369	**0.378**	**0.378**	0.365	0.375

表 12　提案手法による実験の結果（相関）(9 段階)

学習データ	テストデータ	a	b	c	a-b	a-c	b-c	a-b-c
評価者 A	評価者 A	0.514	0.631	0.675	0.585	**0.695**	0.675	0.644
評価者 B	評価者 B	0.491	0.745	0.805	0.633	**0.803**	0.797	0.721
評価者 A	評価者 B	0.654	0.813	0.799	0.836	0.825	0.821	**0.842**
評価者 B	評価者 A	0.579	0.716	0.703	0.729	0.722	0.719	**0.731**

び相関を平均するとそれぞれ 0.771，0.743 で，手動判定に近い結果が得られた．9 段階に関しては，正解率および相関を平均するとそれぞれ 0.430，0.774 で， やはり手動判定に近い結果が得られた．

8. 考　　察

特徴量セット a, b, c をそれぞれ単独で用いた場合は，特徴量セット a（音響特徴量）を用いた場合が最も精度が低かった．入力音声に関する音響的情報が含まれているのが特徴量セット a のみであることを考えると，(ERJ の発声者である)「日本語を母語とする大学生・大学院生相当の学生」という集団に対しては，発声者がどのような発声をしたかにかかわらず，音節の持つコンテキストあるいは原稿の強勢記号から推定した評価結果を提示したほうが，発声の情報のみから推定するよりも精度が高いということになる．すなわち，強勢記号が記載された原稿を渡されて，発声者が「正しい」と思って読み上げた場合，多くの発声者が似たようなパターンで読み上げることになることが予想される．これは逆に，非日本語を母語とする話者の読み上げ音声を対象にした場合は，ミスマッチが起こることも予想される．

2 つの特徴量セットを組み合わせた場合 (a-b, a-c, b-c) の結果については，a-c の組み合わせがおおむね最も高精度であった．音響的情報を加えることで，精度が向上したものと考えられる．一方，b-c の組み合わせについては，特徴量セット b, c をそれぞれ単独で用いた場合と比べて，さほど精度が向上しなかった．特徴量セット c で用いている収録時に参照した原稿に記載されていた強勢記号は，ある音声学者によって付与されたものであるが，そのラベリングの際に，特徴量セット b に含まれるようなコンテキスト情報が用いられたことが関係していると考えられる．すなわち，特徴量セット b から得られる情報と特徴量セット c から得られる情報は相関が高く，両者を組み合わせても，精度向上に寄与しなかったと考えられる．また，全ての特徴量を用いた場合 (a-b-c) は，特徴量セット a-c にくらべて若干精度が下回ったが，偶然なのかどうかは不明である．

おおむね最も高精度であったのは特徴量セット a-c を用いた場合であったが，評価者クローズドの手動判定の精度には及ばなかった。これは各々の評価者が，今回検討した特徴量以外の独自の特徴量に着眼している可能性がある．なお，評価者クローズドの結果と評価者オープンの結果を比べると，一部，評価者オープンのほうがクローズドの場合よりも高精度となって

いる．表 3 および表 4 からも，クローズドの結果が一般に高精度となるはずであるが，逆の傾向を示した結果に関する十分な考察はできていない．

9. おわりに

本研究では ERJ データベースのリズム文の読み上げ音声を評価対象として，識別モデルを用いて高精度に文強勢の自動検出を行うことを目指し，音節ごとに強勢の度合いを 3 段階で自動推定する実験を行った．その過程で，評価対象音声に対して，手動判定による強勢の度合いを 9 段階でラベリングする作業を行ったが，3 段階に丸めて用いた場合と，9 段階のまま使用した場合の 2 通りについて実験を行った．特徴量セットに関して様々な検討を行い，3 段階の場合で，評価者クローズドで最高で 83.1%の正解率を得ることができたが，手動判定の精度には及ばなかった．評価者オープンの場合は最高で 77.1%の正解率を得られ，手動判定の精度に近い結果が得られた．

今後の課題としては，特徴量セットや識別器などの改良による精度向上を考えている．また，手動判定の評価基準をより明確なものとした上で評価者を増やし，より信頼のおける手動判定ラベルデータを構築することを考えている．

文　献

[1] 文部科学省，小学校学習指導要領，2008.

[2] 日本英語検定協会 英語教育研究センター，小学校の外国語活動に関する現状調査 ⟨⟨ 教育委員会対象 ⟩⟩ 調査結果報告，2012.

[3] 日本英語検定協会 英語教育研究センター，小学校の外国語活動に関する現状調査 ⟨⟨ 小学校対象 ⟩⟩ 調査結果報告，2012.

[4] CAI メディア共同開発，英語発音美人 Vol. 1–Vol. 5，2005–2006.

[5] プロンテスト，発音力，2007.

[6] アデュー，発音 PRO，2008.

[7] 学習研究社，えいご三昧 DS，2009.

[8] EnglishCentral, http://www.englishcentral.com/.

[9] 根間弘海，鈴木俊二，英語とリズムをマスターする英語音声学，英宝社，2000.

[10] James L. Hieronymus, “Automatic sentential vowel stress labelling,” *Proc. EUROSPEECH*, 1226–1229, 1989.

[11] Nobuaki Minematsu, Satoshi Kobashikawa, Keikichi Hirose and Donna Erickson, “Acoustic modeling of sentence stress using differential features between syllables for English rhythm learning system development,” *Proc. ICSLP*, 745–748, 2002.

[12] Kazunori Imoto, Yasushi Tsubota, Antonie Raux, Tatsuya Kawahara and Masatake Dantsuji, “Modeling and automatic detection of English sentence stress for computer-assisted English prosody learning system,” *Proc. ICSLP*, 749–752, 2002.

[13] Chaolei Li, Jia Liu and Shanhong Xia, “English sentence stress detection system based on HMM framework,” *Applied Mathematics and Computation*, 185 (2), 759–768, 2007.

[14] 峯松信明，富山義弘，吉本啓，清水克正，中川聖一，壇辻正剛，牧野正三，“英語 CALL 構築を目的とした日本人および米国人による読み上げ英語音声データベースの構築，” 日本教育工学会論文誌，27 (3)，259–272，2003.

[15] Speech Signal Processing Toolkit (SPTK) Version 3.5, http://sp-tk.sourceforge.net/, 2011.

[16] Package ‘kernlab’ Version 0.9-15, http://cran.r-project.org/web/packages/kernlab/index.html, 2012.

[17] A. C. Gimson, *An introduction to the pronunciation of English*, 1962（竹林滋訳，ギムスン 英語音声学入門，金星堂，1983）.

[18] 榎本正嗣，日英語 話し言葉の音声学，玉川大学出版部，2000.

[19] 窪薗晴夫，音声学・音韻論，くろしお出版，1998.