

英語文リズム学習支援を目的とした 文強勢音節のモデル化とその検出

小橋川 哲[†] 峯松 信明^{††} 廣瀬 啓吉^{†††} ドナ・エリクソン^{††††}

[†] 東京大学大学院工学系研究科

^{††} 東京大学大学院情報理工学系研究科

^{†††} 東京大学大学院新領域創成科学研究科

〒 113-0033 東京都文京区本郷 7-3-1

^{††††} 岐阜市立女子短期大学

〒 501-0192 岐阜県岐阜市一日市場北町 7-1

E-mail: {kobashi,mine,hirose}@gavo.t.u-tokyo.ac.jp, erickson@gifu-cwc.ac.jp

あらまし 近年の国際化社会の進展の中で日本人がより自然な英語を話す必要性が増している。正しい英語の発音を身につけるには、強勢音節が等時的に繰り返すという英語特有のリズムを習得する必要がある。本研究では、日本人学習者の為の英語文リズム発音教育システムの構築を目的として、音節単位のモデル化に加え、2つの隣接する音節間での韻律的特徴の相対的差分を用いたものや母音区間のみを用いた文強/弱勢音節のモデル化を提案する。先行研究で提案した手法では、話者クローズ/テキストオープン実験において、約91%の性能が得られた。また他の話者に対して同手法を用いたところ、母国語話者に対しては約8割、日本人話者に対しては約7割の同定率が得られた。さらに、隣接2音節間の韻律的特徴の相対的差分を用いたモデル化により約5%の改良が得られ、母音区間のみモデル化により日本人話者に対して約8%の改良が得られた。
キーワード 英語のリズム、強勢音節、隠れマルコフモデル、英語CAI

Modeling of Stressed Syllables for their Detection in English Sentences to Develop an English Rhythm Learning System

Satoshi KOBASHIKAWA[†], Nobuaki MINEMATSU^{††},
Keikichi HIROSE^{†††}, and Donna ERICKSON^{††††}

[†] Graduate School of Engineering, University of Tokyo

^{††} Graduate School of Information Science and Technology, University of Tokyo

^{†††} Graduate School of Frontier Sciences, University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

^{††††} Gifu City Women's College

7-1, Hitoichiba-kitamachi, Gifu-shi, 501-0192 Japan

E-mail: {kobashi,mine,hirose}@gavo.t.u-tokyo.ac.jp, erickson@gifu-cwc.ac.jp

Abstract Current globalization requires Japanese people to speak English with correct pronunciation. To acquire this ability efficiently, they need to master generating English rhythm with positioning appropriate stressed syllables. With the aim of developing an English rhythm learning system for Japanese learners, in this paper we proposed methods to model stressed/unstressed syllables in English sentences, where relative differences in prosodic features between two adjacent syllables and vowel-based modeling in addition to a syllable-based one were investigated. In experiments, with the previously proposed modeling method, about 91% accuracy was obtained as baseline in the speaker-closed/text-open condition. When using other speakers, approximately 80% and 70% accuracy was obtained respectively with native speakers and Japanese ones. The proposed methods showed that, in the speaker-open/text-open condition, 5% relative improvement was obtained by using the relative differences between two consecutive syllables and 8% improvement was observed for Japanese speakers by using the vowel-based modeling.

Key words English rhythm, stressed syllable, HMM, English CAI

1. はじめに

1.1 英語発音教育の現状

近年、計算機能力の向上に伴い、音声認識・合成などに代表される音声情報処理技術や言語情報処理技術の発展により、それらの技術を応用した語学学習支援システムに関する研究が盛んに行なわれており、PC上のソフトとして市場に流通したり、教育機関等で積極的に導入されている例もある。しかしながら、市販のソフトは単にマルチメディア技術を用いて学習者の学習意欲向上を目的としたものも多い。また音声情報処理技術を用いたものでも音声認識技術を直接的に利用した形態が殆んどであり、その結果分節的特徴に基づく処理系となっている。この場合、リズムやイントネーションなどを伝搬する韻律的特徴は無視された処理形態となっており、語学学習支援システムとしては不十分なものとなる[1]。

現在、教育機関で行なわれている語学教育現場では、[i]と[r]に代表される単音などの音韻の発音指導に比べ、リズムやイントネーション等といった韻律的特徴は、指導の困難さから後回しにされがちである[2]。しかしながら、米語母国語話者が最も重要に感じる発音情報が強勢である[3]というように、韻律的特徴の習得が母国語話者とのコミュニケーションには重要であり、実際、外国語教育法の中には、対象とする言語の音韻の学習に先立ち、リズムやイントネーションといった韻律の学習を重要視するものもある(例えば、Vebo-tonal method[4]など)。実験的にも「韻律的特徴だけが米国人のものを入れ換えられた日本人の英語音声は、日本人の韻律的特徴を使った米国人の音声より(米国人の被験者にとって)自然に聞こえた」[5]との報告例がある。

また、英語には「強勢拍のリズム」と呼ばれる“強勢が等時的に繰り返される”という特徴が存在する[6]。それに対し、日本語は一つ一つのモーラがほぼ等間隔で発音される「モーラ拍のリズム」と呼ばれる特徴を持っている。日本人の英語発音能力は世界的に見て、非常に低いレベルにあると言われているが、こうした言語特有のリズム構造の違いが日本人の英語習得の妨げとなっていると思われる。

そこで我々は、日本人学習者の英語(米語)の文リズム習得を目的とした発音教育システムの構築を目指して、文強勢検出に関する研究を行なっている。

1.2 本研究の背景と目的

先行研究において、峯松らは、孤立発音された英語の単語に対する強勢音節の自動検出技術を構築している[7]。そこでは、音節を強/弱勢のみならず、母音の種類や連接する子音からなる音節構造、単語内位置、前後の音節環境(強/弱勢)などを考慮することで、いくつかの音節カテゴリを定義し、HMMを

用いて音節単位でモデル化を行なっている。この際、日本語のアクセントが声の高さで表現されるのに対し、英語では強/弱勢の差が音の強さ、高さ、長さ、そして母音の音質にまで現れるとの知見[8]から、上記の韻律的特徴に加え、粗い分節的特徴としての低次のケブストラムも特徴量に導入している。

また、筆者らは対象を英語文音声に拡張し、先行研究[7]において構築された英単語強/弱勢音節HMMを用いて、予備的な文強勢検出実験を行なった[9]。さらに、米語文音声試料を用いて強/弱勢音節HMMを、その音節の持つ構造や句内位置に関する属性に基づいてカテゴリ化した後に構築した。構築したモデルを用いて、強/弱勢検出実験を行なったところ、位置や構造を考慮したモデルにおいて約9割の性能が得られた[10]。ただし[10]の目的は、強/弱勢音節のモデル化性能を見ることにあり、[9]と同様、検出対象とする英語文音声としては、米語母国語話者による音声を利用している。また、井本らは同様に音節単位でのHMMを用いて、母国語話者/非母国語話者の音声に対し、高い文強勢検出精度を得ている[11]。

本稿ではまず、[10]で構築した強/弱勢音節HMMを用いて、モデル学習用音声とは異なる母国語話者による音声や日本人話者による音声を対象として強勢検出を行なう。また、[10],[11]では、着眼している音節(以下、当該音節)のみの特徴量を用いてモデル化及び検出を行なっている。しかしながら、強勢の置かれた音節は周囲の音節と比較して異なる特徴を持っており、また周囲の音節に影響を与えることが考えられる。そこで本稿では、周囲の音節との韻律的特徴量に関する差分情報を用いた強勢検出の高精度化を検討する。さらに[10]では、音節構造の複雑化により強勢検出に関する性能の低下が指摘されている。そこで、上記の音節カテゴリを用いて、強勢/弱勢音節中の母音区間のみを用いたHMMを構築し、強勢検出に関する実験的検討についても報告する。

2. 英語文音声試料の収集

市販されている教材では、文中の内容語のみに強勢を置き、強勢を強/弱の2レベルで記述するもの、或いは核強勢と呼ばれる強い強勢も加えて3レベルで記述するものが多い。また特に核強勢に関して、通常句内の最後の内容語に置かれるという「句末原則」[12]が存在する。一方、Ericksonらは、英語教育の観点から英語文に対し、音節、単語、句、文の文構造のレベルに基づき、4段階以上のさらに細かい強勢ラベリングを提案している[13]。このラベリングの特徴として、「多音節単語はレベル2以上の音節を原則的に一つ持つ」ことが挙げられる。本実験の為に収録する英語文に対して、米語母国語話者であるErickson自らが上記の手法に基づき、収録前に強勢ラベリングと句境界のマーキングを行ない、そのラ

ベルを意図して発声した音声を収録した。さらに、収録した音声に対して Erickson 自らが聴取し、ラベルの修正を行なった。これらの文の例を表1に示す(注1)。

表1 英語文音声試料の例 (評価用の一例)

1	1	4								
I'm	amused.									
1	1	3	1	1	4					
I'm	amused	by the	man							
1	1	3	1	1	4					
I'm	amused	by the	man	and his	jokes.					
1	1	3	1	1	2	1	4			
I'm	amused	by the	man	and his	funny	jokes.				
2	1	2	4	1						
Betty	cooks	breakfast.								
2	1	2	1	2	4	1				
Betty	often	cooks	breakfast.							
2	1	2	1	2	3	1	1	2	4	1
Betty	often	cooks	breakfast	before	seven.					

1,2,3,4 = 弱/強勢レベル, | = 句境界

上記のように収集した英語文音声試料 812 文のうち、692 文を学習用に用い (以下, Set-A0-a)、残りの 120 文を評価用に用いる (以下, Set-A0-b)。また、他の話者の音声に対しては、Erickson が聴取して話者の選定及び強勢ラベリング、句境界のマーキングを行ない、これを正解ラベルとした。学習に用いた話者以外の評価用米語母国語話者として、英語学習用教材 [14] 中の男性米語母国語話者 1 名によるリズム学習用英語文における計 75 文 (以下, Set-A1) と TIMIT データベース中の米語母国語話者のうち明瞭かつ教育的な発音をしていると判定された話者のうち男女各 30 名による計 180 文 (以下, Set-A2) を用いた。さらに、日本人話者による英語文音声に対しては、母音挿入などによる検出率の劣化は本研究では別問題として捉え、母音挿入などの少ない比較的英語発音能力が高いと判定された日本人話者男女 5 名ずつによる計 125 文 (以下, Set-J) を評価用音声に用いた。ただし、本研究ではレベル 2 以上の音節を強勢音節、レベル 1 の音節を弱勢音節と定義する。強勢音節に関するさらに細かいモデル化は今後の課題とする。以上、本実験に用いた学習/評価用音声試料を表 2 にまとめる。

表2 学習/評価用音声試料

(1) 学習用音声	
Set-A0-a	米語母国語話者女性 1 名 592 文
(2) 評価用音声	
Set-A0-b	Set-A0-a と同一話者 計 120 文
Set-A1	男性米語母国語話者 計 75 文
Set-A2	米語母国語話者男女各 30 名 計 90 文
Set-J	日本人話者男女各 5 名 計 125 文

(注1) : なお現時点では弱勢がレベル 1、強勢がレベル 2~4 に対応しているが、弱勢に関して schwa とそれ以外で区別することを検討している。

3. 当該音節のみに着目した文強/弱勢音節のモデル化と文強勢検出

3.1 HMM のカテゴリ

先行研究 [7] で構築された強/弱勢音節のモデル化技術を応用する。[7] では、単語の強/弱勢を、音節構造、単語内位置、などにより分類し、各音節グループに対し、モデル化を行なっている。[7] や [9] では母音を、長/短母音のみに分類しているが、本研究では長/二重/短母音の分類を採用する。これに対し、井本らはこの母音の分類に schwa を加えた計 4 種類の分類を採用している [11]。英語の強/弱勢による差がスペクトル情報にまで現れることから、音声認識の結果、当該母音が schwa と判定された音節は弱勢とすることも可能であるが、本研究では、強/弱勢検出において、schwa か否かの判定結果は用いていない。また、前述の「句末原則」[12] から句内の位置が、着眼している音節が強勢か否かに関係すると考えられ、単語内位置の代わりに、句内の位置情報を利用することが予想される。そこで、本研究では、以下の以下 4 つのカテゴリについて実験的に検討した。

- simple 強/弱勢 2 クラスのモデル
- ppos 句内の位置 (句頭/中/末) を考慮した 6 クラスのモデル
- str 音節構造 (中心母音の種類と前後に接続する子音系列) を考慮した 24 クラスのモデル
- ppos.str 句内位置と音節構造を考慮した 72 クラスのモデル

ただし、モデル ppos, ppos_str に関しては、少なくとも句頭、句末が共に定義できる句内音節数が 2 以上の音節のみを選び、モデル化、及び評価を行なった。

3.2 音響分析条件と HMM 学習条件

モデル構築時の音響分析条件は表 4.1, HMM 学習条件は表 4 の通りである。また、文中の無声区間の F_0 は前後の有声区間から線形補間する形で求めた。文頭 (文末) の無声区間は、自動抽出された F_0 データの分布から平均 \bar{F}_0 と標準偏差 σ を求め、有声区間から約平均音節時間長前 (後) に、 F_0 が $\bar{F}_0 - 3\sigma$ の値 (藤崎モデル [15] における基底周波数 F_b に相当) にあるものと仮定して線形補間を行ない、全体に渡るスムージングを施した。さらに全発声中の平均 F_0 を求め、対数化した後、この平均 \bar{F}_0 を引いて正規化したものを最終的な F_0 パターンとした。パワーに関する正規化も同様である。

3.3 音節境界の検出

本研究では、学習者は提示された文を発声する学習形態を想定している。この場合、文強勢検出の際に、学習者が意図した文 (発音記号列) や文中の意図した音節数を既知情報として利用できる。[7] では、入

表3 音響分析条件

サンプリング	12kHz/16bit
分析法	14次 LPC 分析
フレーム幅	21.3 msec (256 samples)
フレーム周期	8.0 msec (96 samples)
ピッチ抽出	遅れ時間幅に比例したフレーム幅 を使用した自己相関法
抽出周期	8.0 msec (96 samples)

表4 HMM 学習条件

特徴量	$F_0 + \Delta F_0$, POW+ Δ POW, LPC-MCEP(1~4)+ Δ CEP (合計 12 次元)
topology	6 状態 4 分布, left-to-right 継続時間長制御
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

力音声に対して不特定話者用英語音韻モデルを用いていたが、本研究では音節境界検出精度を高める為、Set-A0-a/b の音声に関しては、学習/評価用話者に適応化した英語音韻モデルによる forced alignment を行ない、得られた音韻境界から [16] により音節境界を求め、切り出した音節を用いて、モデル学習/評価を行なった。その他の評価用音声に対しては、[7] と同様に不特定話者用英語音韻モデルを用いて音節境界を求め、音節への切り出しを行なっている。

3.4 文強勢検出

第3.3節により切り出された音節は、その音節の持つ属性を基に照合対象の強勢/弱勢音節モデルを選択し、音節単位での同定結果が算出される。これらの算出結果を接続することで、文全体での強勢パターンとしての同定結果が得られる。さらに、第2節で述べたラベリングでは、「多音節単語内の強勢音節数は1である」という制約を加えることができる。

3.5 結果と考察

図1に音節単位での同定率を示す。図2に前述した単語内強勢音節数の制約条件を加えて算出した音節単位での同定率を示す。モデル学習用音声 (Set-A0-a) と同一話者の音声 Set-A0-b に関して議論すると、カテゴリ数の増加に伴い性能の向上が見られる。また、単語内強勢音節数の制約により、各モデルで約3%程度の性能向上が見られ、最大で約91%の同定率を得ることができた。

一方、他の母国語話者 (Set-A1/A2) についてみると、句内位置情報を付加することで同定率に向上が見られる。しかしながら、音節構造を導入すると同定率が低下してしまう。この原因として、構築した強/弱勢音節モデルが話者 Set-A0 の個性を含んでいる可能性が考えられ、また forced alignment の際、不特定話者用英語音韻モデルを用いている為、音節切り出しの精度が低下した為とも考えられる。ほとんどのモデルにおいて、制約条件の付加により約3%の同定率の向上が見られ、母国語話者に対しては約8割の同定率を得ることができた。Set-A1 音声に対して、制約条件による性能向上が見られないのは、Set-A1

音声が比較的平易な文であり、単音節単語の占める割合が多かったことに起因しているものと思われる。

それに対し、学習対象である日本人の音声に対しては、同様に句内位置情報の付加により同定率が向上しており、音節構造情報の付加によっても同定率が向上している。また、制約条件により各モデルで同定率に1%以上の向上が見られた。その結果、日本人話者の音声に対して最大で約73%の同定率を得ることができた。

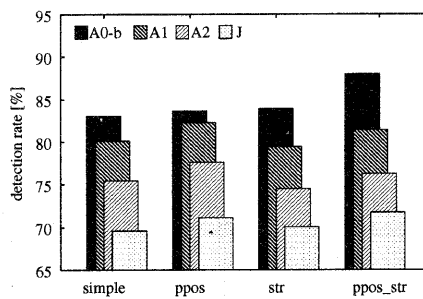


図1 当該音節のみを用いた文強勢検出 (制約無)

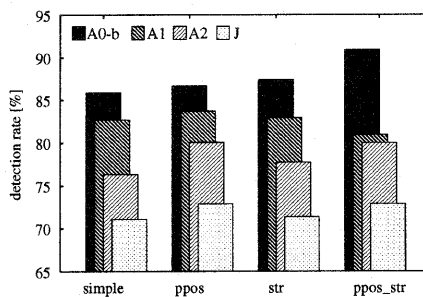


図2 当該音節のみを用いた文強勢検出 (制約有)

4. 隣接音節との差分情報を用いた文強/弱勢音節のモデル化と文強勢検出

第3節では、当該音節のみの特徴量を用いて検出を行なっている。一般に強勢が置かれた音節は周囲の音節と比較してより“目立つ”存在となる。そこで、隣接する2音節の強さと高さに関する特徴量の差分情報を用いた文強勢検出手法を検討する。

4.1 差分情報の利用とHMMのカテゴリ

隣接2音節単位の強/弱勢パターンとして、強強/弱弱/強弱/弱強の4つのが考えられる。この時パワーパターンと F_0 パターンに関して、子音による変動を無視すれば、およそ、強弱では右下がり、弱強では右上がり、強強/弱弱では傾きが小さい水平に近いパターンとなることが予想される。さらに、強強と弱弱では、長さに関して大きな差異があると思われる。そこで、この4パターンの相対的な動きを基にしたモデル化を考える。2音節単位でのモデル化を行なう場合、前述の手法と同様に各音節の持つ属性に従ったモデル化を行なうとカテゴリ数が膨大となり、モデル構築の為のデータ量の不足が予想される。そこ

で、当該音節の前後に隣接する音節の特徴量との差分情報を反映し、当該音節のみを対象にしたモデル化を行なう。句内において、当該音節の前後に接続する音節が存在するもの（いわゆる句頭/末を除いたもの、すなわち句中音節）を対象とする。まず、当該音節とその前に接続する音節（以下、前接音節）の2音節における F_0 を対数化したのち平均値 \bar{F}_0 を求め、この \bar{F}_0 を引く形で正規化を行った。パワーに関しても同様に2音節単位で正規化を行った。これにより前接音節との2音節単位で正規化を行った当該音節のデータ（以下、前接データ）が得られる。また、当該音節とその後に接続する音節（以下、後接音節）との2音節単位で正規化を行った当該音節のデータ（以下、後接データ）も得られる。以上により、句中音節の音節数に対して2倍の前/後接データが得られる。ここで、当該音節と前/後接音節の強/弱勢の組を（当該音節、前/後接音節）と表現する。接続音節が前接であるか後接を無視すると、以下の4つの強/弱勢に関するカテゴリを定義することができる。すなわち、Ss:(強勢, 強勢), Ww:(弱勢, 弱勢), Sw:(強勢, 弱勢), Ws:(弱勢, 強勢) である。これにさらに、当該音節の音節構造などの属性を加えることによって、カテゴリ数を増やすことができる。これら隣接する音節の特徴量を用いて正規化された音節を用いて構築されたモデルをまとめて隣接正規化音節モデルとする。このモデルでは、以下の3つのカテゴリについて実験的に検討した。隣接正規化音節モデルでは句中音節のみを対象としている為、位置情報は既に付加されている。そこで、以下のカテゴリ名は、第3.節で述べた当該音節のみを見て構築したモデル（以下、当該音節モデル）に対応するように付けている。

ppos	Ss/Ww/Sw/Ws の4クラスのモデル
ppos_vow	音節中の母音構造（長/二重/短母音）を考慮した12クラスのモデル
ppos_str	音節構造（中心母音の種類と前後に接続する子音系列）を考慮した48クラスのモデル

4.2 文強勢検出

第4.1節により正規化された前/後接データは、その音節の持つ属性を基に照合対象の隣接正規化音節モデルを選択し、前/後節音節それぞれに対しての強/弱勢に関する同定結果が算出される。本研究では、前/後節データの両同定結果が一致する場合についてのみ議論する。

4.3 結果と考察

図3に前/後接データの両同定結果が一致する場合の音節に対する隣接正規化音節モデルと当該音節モデルとの同定結果を示す。

母国語話者の音声に関しては、全モデルに対して隣接正規化音節モデルの性能が当該音節モデルに勝っ

ている。特に、Set-A1 音声では最大約5%、Set-A0-b 音声では最大約6%以上の性能向上が見られる。これは両音声は英語の文リズムの発音教育を念頭に置いて発声された文であり、当該音節と前/後接音節との間に明確な差を持つ音声であるからであると考えられる。また、この両者間の隣接正規化音節モデルにおける性能差は小さく、このモデルが話者性がある程度正規化され、リズムに特化したモデルであることが示唆される。

また、日本人音声に関しても、モデル ppos/ppos_vow については約2%性能向上が見られる。一方、ppos_str に関して性能向上が見られないのは、当該音節モデルと同様の理由によるものと思われる。

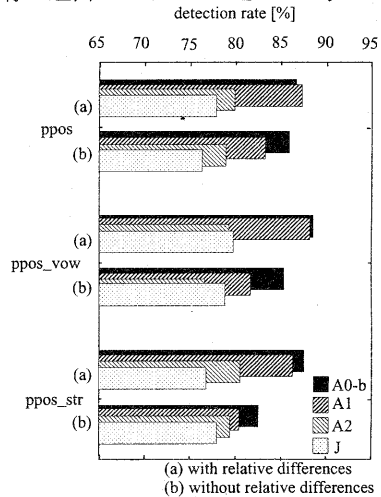


図3 当該/隣接正規化音節モデルの比較

(a) 隣接音節正規化モデル, (b) 当該音節モデル

5. 音節カテゴリを用いた母音区間のみによる文強/弱勢のモデル化と文強勢検出

5.1 母音区間のみを用いた強勢検出

図1, 図2において、Set-A1/A2 音声に対する str モデルの同定率は simple より低く、また前節で ppos_str より ppos_vow の方が高い性能を示している。その原因の一つとして音節構造の複雑化により子音を含めた音節切り出しの精度が問題であると思われる。しかし適応化された英語音韻モデルを用いて高精度な音節切り出しがされている Set-A0 の音声に対しても、音節構造の複雑化により強勢検出に関する性能の低下が指摘されており [10]、音節の持つ構造により性能が変化する強勢検出器は望ましくない。そこで音節に切り出すのではなく、母音単位での特徴量に対する強勢検出を行なうことが考えられる。しかしながら、[6]をはじめとする多くの英語音声学に関する文献では、強/弱勢は音節単位で議論されている。また一方で、[17]のように母音単位で強勢を議論している例も存在する。そこで本実験で

は、特徴量としては母音単位で切り出したデータを用い、モデル構築の際のカテゴリは、その母音を中心とする音節の属性を用いて強勢検出を行なう。従って、HMMのカテゴリに関しては、第3.1節と同じ4カテゴリにおいて実験を行なった。

5.2 結果と考察

強/弱勢母音のモデル化による文強勢検出結果(制約無/有)を図4, 図5に示す。この結果と図1, 図2を比較すると、母国語話者に関してはやや性能低下が見られるもののほぼ同程度の性能が見られる。一方、日本人話者に対する結果では、当該音節モデルを基準として最大で約8%の性能向上が見られ、約76%の同定率を得ることができた。これは日本人話者が母国語話者に比べ、意図した音節構造通りの発声に困難である為と考えられる。また全てのモデルにおいて、単語内強勢音節数制約の効果が現れている。さらに実装のしやすさを考えても有効な強勢検出技術であると言える。またSet-A0-b/A1に関しては音節構造の情報を用いることによる性能向上があまり見られない。母音区間のみをモデル化する際、当該母音を含む音節構造を反映することは必ずしも適切ではないことが示唆されるが、これに関しては実験的検討が必要である。

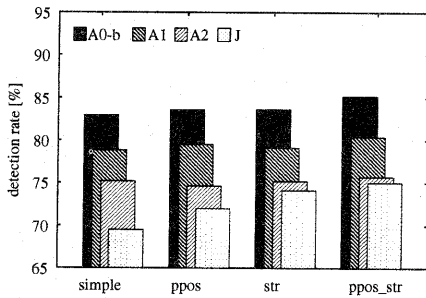


図4 母音区間を用いた文強勢検出(制約無)

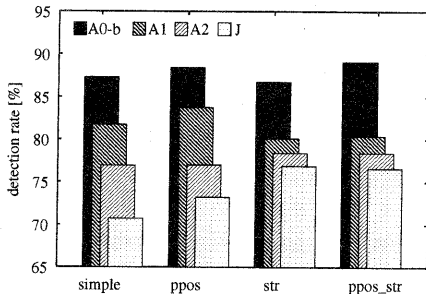


図5 母音区間を用いた文強勢検出(制約有)

6. まとめ

英語音声から各音節の持つ属性に従い、当該音節のみを用いて文強/弱勢音節モデルを構築した。構築したモデルを用いて、文強/弱勢検出を行なったところ、句内位置情報と音節構造を考慮することで、学習時と同一話者による評価音声に対して約91%の性

能を得ることができた。また、他の英語母国語話者の音声に対して約8割の同定率を得ることができ、対象としている日本人話者の音声に対して、約73%の結果を得ることができた。また隣接音節との差分情報を利用した強勢検出技術を構築し、一部のデータに関して、当該音節のみを用いたモデルと比べ約5%の同定率の向上を計ることができた、と同時に話者性を回避する可能性を示すことができた。さらに、音節カテゴリを用いた文強/弱勢母音モデルを構築し、母音単位での文強/弱勢検出を行なったところ、日本人話者に対して当該音節のみのモデルと比較して約8%の性能向上を計ることができ、結果として約76%となる同定率を得ることができた。

今後は、日本人英語の特性などを考慮するなど、本研究で得られた強勢検出技術を実際の英語文リズム発音学習システムへどのように利用していくべきかが課題となる。

文 献

- [1] <http://www.slp.tutics.tut.ac.jp/CALLsoft/>.
- [2] 斉藤明子, 長井克己. VT法を用いた日本語音声教育の実践報告. 第1回日本語音声教育方法研究会, 1999.
- [3] 河合剛, 石田朗. 日本人の英語の発音評価の信頼性に関する実験的研究. 信学技報, Vol. ET95-44, pp. 89-96, 06 1995.
- [4] <http://www.kanto-gakuen.ac.jp/verbo/verbo.htm>.
- [5] A. Weibel. Prosody and speech perception. *PhD Thesis, Carnegie-Mellon University*, 1986.
- [6] 竹林滋. 英語音声学. 研究社, 1996.
- [7] 峯松信明, 藤澤友紀子, 中川聖一. HMMを用いた英単語音声からの強勢音節の自動検出とそれに基づく発音能力の韻律的評定. 電子情報通信学会誌 D-II, Vol. J82-D-II, No. 11, pp. 1865-1876, 11 1999.
- [8] A.C.Gimson 著, 竹林 滋訳. ギムスン 英語音声学入門. 金星堂, 1990.
- [9] 峯松信明, 小橋川哲, 廣瀬啓吉. 語学学習支援を目的とした英語音声からのリズム同定に関する実験的検討. 日本音響学会講演論文集, 3-3-6, pp. 331-332, 春季 2001.
- [10] 小橋川哲, 峯松信明, 廣瀬啓吉. 英語文リズムの学習支援を目的とした強勢検出. 日本音響学会講演論文集, 1-Q-4, pp. 147-148, 秋季 2001.
- [11] 井本和範, 壇辻正剛, 河原達也. 日本人話者の英語文強勢誤りの自動検出. 日本音響学会講演論文集, 3-7-2, pp. 343-344, 秋季 2001.
- [12] 小川直樹. 理屈で分かる英語の発音 Speak Naturally: A guide to English Rhythm and Intonation. NOVA, 2000.
- [13] ドナ・エリクソン. 英語における会話時の顎の動きとリズム. 日本音響学会聴覚研究会資料 H-98-59 電子情報通信学会技術研究報告 SP98-41, pp. 1-8, 7月 1998.
- [14] 深澤俊昭. 英語の発音パーフェクト学習辞典 A Guide to Perfect English Pronunciation. アルク, 2000.
- [15] H. Fujisaki and S. Nagashima. A model for synthesis of pitch contours of connected speech. *Annual Report of Engineering Research Institute, University of Tokyo*, Vol. 28, pp. 50-60, 1969.
- [16] <http://www.nist.gov/speech/tools/tsylb2-11tarZ.htm>.
- [17] Chao Wang and Stephanie Seneff. Lexical Stress Modeling for Improved Speech Recognition of Spontaneous Telephone Speech in the JUPITER Domain. *Eurospeech 2001-Scandinavia*, Vol. 4, pp. 2761-2764, 2001.