論　　文 43.72.Ne

英語韻律発音学習支援システムのための英語文強勢のモデル化と自動検出*

井本和範*1　坪田　康*1　河原達也*1　壇辻正剛*1,*2

[要旨] 外国語による意思疎通において韻律要素の役割は大きく，その習得は発音学習における重要な課題の一つである。本研究では日本人学習者の英語文発声を韻律的な側面，特に文強勢に着目して発音評価・教示する方法を提案する。文強勢を自動検出するために日本人の誤り傾向を分類し，音節構造（母音種を含む）・句内位置のカテゴリごとに 3 段階の強勢 HMM を構成する。これにより種々の発音誤りの影響を考慮して強勢の有無を判定すると共に，誤り原因を学習者に提示することができる。更に判別分析により得られた母語話者の知覚傾向を各音響特徴量の重みづけとして，多段階識別により効果的に反映する手法も提案する。米語母語話者と日本人学習者に対する評価実験の結果，おのおの 95.1%，84.1%の強勢識別精度を得た。

キーワード　韻律，CALL，文強勢，HMM，多段階識別
Prosody, CALL, Sentence stress, HMM, Multi-stage discrimination

1. はじめに

外国人と意思疎通を行う場面が増加する現在，英会話能力は必要不可欠な技能となりつつあるが，その習得のために克服すべき発音上の問題点は少なくない。日本人が英会話を習得する際における問題の一つとして，強勢，リズム，イントネーションなどの韻律要素が挙げられる。教育現場においても，指導が難しく習得も容易でないなどの理由から，韻律学習を後回しにする傾向がある[1]。しかし韻律要素の誤りは，意思疎通を妨げる致命的な発音誤りとなる可能性がある。実際，/l/と/r/などの音韻誤りには，それが首尾一貫していれば母語話者はすぐに順応できるが，強勢位置の誤りには容易に順応できないと言われている[2]。更に日本人英語の了解度を下げる最大の要因がイントネーションやリズムにあることが報告されている[3]。また，韻律要素は，特定の単語を強調したり，話者の意図を伝達するため，意思疎通の上で非常に重要な機能を担う[4]。発音学習がコミュニケーションの実現を目的とする以上，個々の子音や母音の音韻要素のみでなく，韻律要素の学習が重要である。

一方，音声認識に代表される音声情報処理技術を応用して，発音指導の面から非母語話者の言語習得を支援する計算機支援型言語学習（Computer-Assisted Language Learning: CALL）の研究が近年盛んに行われている[5-9]。CALL システムは，単調な反復学習に伴う作業を自動化することで効率的な学習を実現し，更に映像や音声を融合することで対話的な学習環境を実現できる[10]。しかし従来の CALL システムは，音声認識技術による音韻学習や，対話的な学習環境の再現に焦点を当てたものが多く，韻律的な側面からの評価や教示に関する検討は十分になされていない。

非母語話者の文発声の発音を評価する際には，(1) 文全体の定量的な発音スコアの評定，(2) 個別（韻律の場合はリズムやイントネーションなど）の発音誤りの指摘という二つのアプローチが考えられる。教師話者と比較して F_0 やパワーの差分距離をスコア化する方法[11]は前者に，文強勢の位置が異なる音節や F_0 曲線の概形が異なる文末句を検出する方法[12]は後者に対応する。発音教示において，前者は発音習得レベルの把握に，後者は矯正すべき誤り個所の同定に重要な役割を果たす[9]。しかし具体的な発音矯正には，誤り同定の重要性がより高いと考えられる。

誤り個所を同定する際には，学習者が母国語の発音体系を参照する可能性に着目することが重要である。母国語と学習言語の両方の発音体系を用いて誤り規則

---

* Modeling and automatic detection of English sentence stress for computer-assisted English prosody learning system,
by Kazunori Imoto, Yasushi Tsubota, Tatsuya Kawahara and Masatake Dantsuji.

*1 京都大学大学院情報学研究科
*2 京都大学学術情報メディアセンター
(問合先：河原達也　〒606-8501 京都市左京区吉田本町　京都大学大学院情報学研究科)

(2002 年 3 月 18 日受付, 2002 年 11 月 13 日採録決定)

を記述すれば，母国語の影響を受けた学習者の誤りを頑健に検出できる[9]。実際に単語発声では，日本語と英語の言語構造の差違を利用して単語強勢を検出し，音響的な差違に着目した教示を行う手法が提案されている[13]。しかし文発声の評価に適用する場合，イントネーションなどの文発声特有の現象に対応する必要がある。更に，フレージングや機能語の弱形といった単語発声では影響の小さい他の発音要素に関しても，文強勢に与える影響を考慮する必要がある。

本研究では，日本人の英語を対象に，文強勢に注目した発音評価法に焦点を当てる。発音評価は，HMMによって文強勢音節の有無を判定することで行う。その際，母語話者の単語発声を対象とする先行研究[14-17]と異なり，日本語と英語に存在する言語構造の差違，文発声特有の現象及び他の発音要素の影響を考慮して，文強勢に影響のある誤りを分類する。誤りに応じてカテゴリ化された文強勢音節のHMMを作成することで，日本人英語の文発声に対して，文強勢を高精度に自動検出する方法を提案する。更に，種々の要因を考慮することで増加した強勢の種類を，多段階で効果的に識別する手法も提案する。本手法の有効性を米語母語話者と日本人の両音声を用いて検証する。

2. 日本人の文強勢誤り傾向

本章では，日本語と英語の言語構造の差違に着目して文強勢における発音上の問題点を分類し，日本人の誤り傾向に応じて作成したカテゴリについて述べる。

2.1 英語文強勢

英語の強勢は音節を基本単位とした言語情報であり，様々な音響的特徴の相対的な差違によって表現される。日本語ではアクセントが類似の言語的役割を果たすが，基本単位となる音節構造や音響的特徴が英語と日本語では大きく異なる。そのため日本人が英語を発声する場合には，日本語アクセントの影響を受けて，様々な発音誤りが観測されることが多い。以降，(1)英語と日本語の構造的差異，(2)英語強勢と日本語アクセントの音響的差異，(3)文強勢と単語強勢の差違をおのおの比較して，それらに起因する発音上の問題について考察する。

2.1.1 英語と日本語の構造的差違

言語音を構成する最小単位は音素であるが，発声の最小単位は英語では音節，日本語ではモーラとなる。英語強勢では音節が，日本語アクセントではモーラが各々の基本単位となる。それらの比較を**表-1**に示す。英語では，子音連結や単語末での子音が許されるために，音節の種類は日本語に比べてはるかに多く，1,000以上も存在する。一方日本人の英語には，開音節構造を基本とする日本語の影響を受けて，しばしば誤った母音が挿入される。

表-1 英語強勢と日本語アクセントの差違

英語強勢	日本語アクセント
基本構造	
音節	モーラ
—母音(V)を核に前後に0以上の子音(C)が連結(V, CV, …, CCCV-CCCC)	—基本はV，CV他に特殊拍(促音，撥音，長音)
—開/閉音節の両方が存在	—基本的に開音節
音響的特徴	
強弱アクセント	高低アクセント
—声の強さ，声の高さ，音の持続時間，母音音質が複雑に関与	—基本的には声の高さのみが関与

2.1.2 強勢とアクセントの音響的差違

日本語は高低アクセント，英語は強弱アクセントと記述されることが多い。これは日本語のアクセントが主に声の高さの変化で表現されるためである。一方，英語では必ずしも強さの変化のみで強勢が表現されるわけではない。英語の強勢には声の強さに加えて，声の高さ，持続時間，母音音質が関与する[18]。このように関与する特徴が多いため，英語強勢の発声には複雑な音響的変化が伴う。

2.1.3 文強勢と単語強勢の差違

孤立単語発声では，強勢を持つ音節の位置は単語と品詞によって一意に決定される。一方，文発声では，単語発声と比較して強勢の位置が変化する。前後の文脈や特別な強調を持たない単独の文を発声する場合，基本的には名詞や動詞などの内容語は強勢を受けるが，冠詞や前置詞などの機能語は強勢を受けにくいという規則（ノーマルストレスと呼ばれる）が存在する[1]。そこに強勢が等間隔に出現しようとする作用が働き，強勢位置の移動が起こる。更に，話者の意図や文脈などの影響を受けて強勢位置が変化するため，自由発話において適切な強勢位置を予測するのは容易でない。しかし語学教育システムでは，スキットなどで文脈をあらかじめ定めることで発話文の正解強勢位置を推定することができる。

上記に加えて文発声では，イントネーションや句切りの影響を考慮する必要がある。自然発声のピッチは，発話冒頭で急激に上昇し，その後徐々に下降する山型の曲線を描くため，発話の句切りごとにピッチ曲線のまとまり（音調単位と呼ばれる）が形成される。単語発声では，音調単位を一つと仮定して問題はないが，文が長くなればその数は増加する。特に不適切に

単語間を区切る可能性のある非母語話者の発声では，音調単位の数が増加することが予測される。

2.2 日本人の誤り傾向

前節で述べた言語構造の比較から，日本人が発声する英語に含まれる文強勢の誤り原因を分類する。

2.2.1 ピッチアクセントの影響

英語強勢を発声する際に，日本語アクセントの発声方法，すなわち声の高低変化を顕著に用いる誤りを引き起こすことが知られている[18]。

2.2.2 発音誤りによる音節構造の変化

日本人話者の英語に頻繁に生じる母音挿入は，音節構造の変化や音節数の増加を招き，結果として強勢の誤りを引き起こす。また，機能語の多くには，強勢の有無に対応して発音そのものが変化する強形・弱形が存在する。弱形における音節構造の変化や母音の曖昧化によって，相対的に他の強勢音節を際立たせる。この強形・弱形も日本人が苦手とする発音要素の一つで，不必要に強勢音節を増加させる傾向がある。

2.2.3 発声句切りの影響

発音を十分に習得していない単語や，名詞を限定する複数の形容詞を，統語的な構造を考慮しないで句に分割する例がこれにあたる。不適切な位置で文を区切れば音調群が増加し，ピッチ変化に伴う強勢音節が必要以上に増加する。

2.3 誤り傾向に応じた文強勢のカテゴリ化

前節で述べた日本人英語の特徴や文発声特有の現象に対応するために，これらの誤り傾向に応じた詳細な音節カテゴリを作成する。

2.3.1 強勢の段階

英語の強勢は，強勢と無強勢の2値ではなく複数の段階が存在し，音声学的には第1強勢，第2強勢，第3強勢と無強勢の4段階も存在すると言われる。しかし強勢を単に多段階の体系という観点から考えるのは，音韻的事実と合致しないため適切でないとの意見もある[19]。特に発音学習を目的とした場合，一貫した音韻的差違がないものを学習・教示するのは適当でない。そこで本研究では音韻的事実に注目した文献[1]における強勢の定義を採用する。それによると，ピッチ変化を帯びる核音調を持つ強勢を第1強勢(Primary Stress: PS) とし，音調群に基本的に一つ存在する。また，第1強勢以外の強勢を第2強勢(Secondary Stress: SS) とし，強勢を持たない無強勢（No Stress: NS）と合わせて3段階で表現する。この定義ではピッチアクセントの影響による誤りを，PSとSSの違いとして検出できる。このように強勢の種類によって，ピッチアクセントの影響を考慮する。

2.3.2 音節構造の導入

音節構造と強勢の有無の間には一定の相関があり，CVCのような複雑な構造を持つ音節ほど強勢となる可能性が高くなることが指摘されている[20]。しかしながら，英語の音節の種類は1,000以上にも及ぶため，すべての音節と強勢の関係をモデル化することは事実上不可能である。そこで本研究では音節構造をV，CV，VC，CVCの4種類にカテゴリ化する。ここでVは母音を，Cは子音系列を表す。同様に，母音の種類と強勢の間にも相関があることを考慮して，曖昧母音（Vx），短母音（Vs），長母音（Vl），二重母音（Vd）の母音種も導入する。すなわち構造（4種）と母音種（4種）を組み合わせた16種類のカテゴリを用意する。このような詳細な分類は，音節構造と強勢の対応を高精度にモデル化できるだけでなく，母音挿入や機能語の弱形発音といった，日本人英語の音節構造誤りへの柔軟な対応を可能とする。

2.3.3 句内位置の導入

2.1.3節で述べたように，句の先頭と末尾部分ではピッチ曲線の振る舞いが大きく異なる。従って，各音節の句内における位置（以降では句内位置と呼ぶ）によって韻律パターンは異なることが予測される。本研究では，句頭（H），句中（M），句末（T）の句内位置によって音節を分類する。これは同時に，不適切に単語を独立した句として発声する誤りへの対応を可能にする。

なお，単語発声の強勢検出を対象としていた先行研究[13]で用いられているカテゴリと比較して，本研究では文発声を対象としているため，(1)強勢を3段階で分類，(2)単語内位置ではなく句内位置で分類，(3)音節構造を規定する母音種として，短母音(Vs)・長母音（Vl）だけでなく，二重母音（Vd）と曖昧母音（Vx）も追加している，などの違いがある。

3. 文強勢の自動検出

3.1 HMMによる文強勢のモデル化

本研究では，HMMによって文強勢音節の自動検出を行う。以降では作成した文強勢HMMの構成及び特徴抽出方法について説明する。

3.1.1 文強勢HMMの構成

学習した文強勢HMMの構成と音響分析条件を**表-2**に示す。最小単語対の強勢をHMMで自動検出する先行研究[16]では，F_0・パワーに加えて4次元までのLPC係数を特徴量としている。これは低次LPC係数によって粗いスペクトル情報，すなわち母音の音質を強勢検出に利用するものである。その他の多くの先行研究でもF_0・パワー・音素持続長・LPC係数の

表-2 文強勢 HMM の構成

特徴量	log(F_0)，log(Power)，4 次元 MFCC と各 1 次/2 次微係数（計 18 次元）
分析フレーム	フレーム周期 10 ms フレーム幅 25 ms
HMM の構成	3 状態 HMM（left-to-right） 混合数 8，対角共分散行列 3 ストリーム（F_0，パワー，MFCC）

複数の組み合わせで特徴ベクトルを構成している。これは表-1 に示す英語強勢の音響的特徴を反映した結果である。本研究では，F_0・パワー・1〜4 次までの MFCC，及び各々の 1 次微係数と 2 次微係数を含む合計 18 次元の特徴ベクトルを用いる。2 次微係数は予備的な実験によって導入を決めた。また，各特徴量は互いに無相関と仮定して，F_0・パワー・MFCC ごとに別のストリームを構成した。音素持続長に関しても先行研究[13]のようにモデル化するのが望ましいが，本研究では未実装であり今後の課題である。HMM は 3 状態の left-to-right 型とし，飛び越しを許さない状態遷移行列とした。また，混合数に関しては，予備実験の結果から 8 とした。

2.3 節で定義したカテゴリに基づいて，3 段階の文強勢それぞれに対して，16 種類の音節構造と 3 種類の句内位置の組み合わせごとに HMM を作成する。すなわち最も詳細な場合で，3（段階）×16（音節構造）×3（句内位置）の 144 種類の HMM が作成される。

3.1.2 音響特徴量の抽出

F_0 とパワーの抽出方法について述べる。F_0 の抽出は PARCOR 分析によって得られる予測残差波形の自己相関を用いて行った。有声区間では，自己相関値が極大となる複数の F_0 候補を，連続性を考慮したビーム探索によって絞り込むことで値を決定した。無声区間では，前後の有声区間の F_0 抽出結果に基づいて 3 次元スプライン補間を行った。文発声においては，途中で不規則に休止が挿入されることがあるため，（無声音と無音を区別して）補間を行うべき時間長を考慮する必要がある。ここでは 300 ms 以内であれば補間処理を行うこととした。F_0 とパワーは 10 ms ごとに抽出し，それぞれを対数化し，各文発声ごとに平均値を 0 とする正規化処理を行い，最終的な特徴量とした。

3.2 HMM による文強勢の検出方法

文強勢を自動検出する処理の概要を図-1 に示す。処理は大きく三つの部分から構成される。

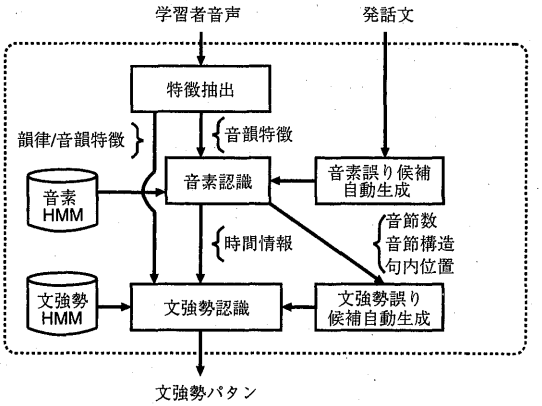


図-1 文強勢誤り自動検出方法の概要

3.2.1 音節区分化と音節構造・句単位の同定

表-2 に述べた韻律を中心とした特徴量のみで，音素列と音声の時間的な対応付けを行うことは難しいので，事前に音声認識システムを用いて音節区分化を行う。基本的には，発音音素列に基づいて音素 HMM を連結したモデルによるビタビアライメントによって音節境界の決定を行うが，非母語話者の発声には，音素発音誤りや不規則な休止が含まれるという問題がある。本研究では，母語話者のみの音声から学習した英語音響モデルを用いており，辞書表記の音素列を用いた強制アラインメントでは十分な精度が得られないと考えられる。

そこで，英語音声学などの知見に基づいて誤り規則を作成し，発声テキストに対して音素発音誤りを含む認識候補を動的に生成することで，区分化精度の改善を図る[21]。これにより，学習者の発声が辞書表記と異なったり，不適切な文の区切り方がなされても，音節構造と句境界の頑健な検出を行う。具体的には，音素誤りパターン候補の中に母音挿入や強形と弱形の違いといった音素列を含むことで音節構造の検出を行う。また，すべての単語間に休止が挿入される可能性を認識ネットワークに記述することで句境界の検出を行う。句境界の検出により，各音節の句内における位置が決定される。以上の処理によって，音節数，音節構造，句内位置，及び各音節と音声の時間的対応が得られる。

3.2.2 文強勢誤り候補の自動生成

音声認識の結果に基づいて文強勢の認識ネットワークを作成する。最も詳細なモデルでは 144 種類の HMM を用いるが，検出された音節構造と句内位置を音節ごとに固定することで，各音節がどの強勢種（PS, SS, NS）に属するかを判定すればよい。つまり，この段階で音節数 N に対して，認識候補は 3^N と

なる。

ただし，多音節単語を単独発声する場合に存在する複数の強勢音節は，最も強い音節を除いて文中では強勢を失う可能性が高い[1]。実際，(米語母語話者はもちろんのこと) 4.1節で述べる日本人の発声した英語に対する専門家のラベルを観察しても，1単語中に強勢音節が複数存在する場合はなかった。従って強勢音節の数は単語中に一つであると仮定し，認識候補に制約を課す。更に，PS (核強勢) は音調群 (=句単位) ごとに最大一つ存在すると仮定する。

図-2 に文強勢の誤り候補の生成例を示す。

3.2.3 文強勢 HMM による誤り検出

作成した文強勢の認識ネットワークと文強勢 HMM を用いて，強勢パターンの識別を行う。この強勢 HMM による識別は，まず音声認識システムで同定された単語境界を用いて単語単位で可能な強勢パターンを用いて行う。その後で句単位で PS (核強勢) は高々一つという制約を適用して，PS が複数出現している場合は尤度が低い方を SS (第2強勢) に修正する操作を行う。文や句のような長い単位に強勢 HMM をそのまま適用しても，正しい音節との対応付けの信頼性が十分でない可能性があるため，区分化情報をできるだけ利用することとした。

辞書から作成した正解強勢列と認識された強勢列を比較して，強勢種が異なる場合にその個所を誤りと判定する。更に，音節構造や句内位置が正解と異なる場合には，それらが誤りを引き起した可能性があることを学習者に提示する。

3.3 強勢音節の多段階識別の導入

3.3.1 HMM のストリーム重み

本研究で用いる強勢 HMM の Viterbi スコア $f(i, t)$ は次のように表現される。ただし，i, j は状態，t は時刻を表す。

$$f(i, t)=\max_j[f(j, t-1)\cdot a_{ji}\cdot b_i(y_t)] \qquad (1)$$

ここで，a_{ji} は状態 j から i への状態遷移確率を表し，$b_i(y_t)$ は状態 i で特徴ベクトル y_t を出力する出力確率を表す。F_0・パワー・MFCC を別ストリームでモデル化しているので出力確率は以下のように記述できる。

$$b_i(y_t)=\{b_i^1(y_t^1)\}^{w1}\cdot\{b_i^2(y_t^2)\}^{w2}\cdot\{b_i^3(y_t^3)\}^{w3} \qquad (2)$$

ここで，$b_i^x(y_t^x)$ は各ストリームの出力確率であり，x は番号順に F_0，パワー，MFCC を表す。また w_x は，各特徴量に対する重みを表し，その値が大きいほど，対応する特徴量が文強勢の識別において重要であることを示す。これは母語話者が知覚する際にどれを重視するか (知覚傾向) に対応する。ベースラインではすべての重みを1にしている。

3.3.2 多段階識別

各強勢 (PS, SS, NS) の特性を考えると，各々の識別に重要となる音響的特徴が異なる可能性が高い。すなわち，PS は核音調を含む強勢に対応するラベルであるため，PS を他の強勢種と区別する場合には，F_0 の重要性が高くなることが予想される。一方，SS と NS を識別するためには他の特徴量が重要である可能性がある。このように識別対象に応じて，利用する特徴量の重みを変えることを考える[22]。

そのために各強勢の識別を2段階で行う。処理の流れを**図-3** に示す。第1段階では，文強勢 (PS 又は SS，以降は ST) か無強勢 (NS) かの判定を行う。ここで示す ST モデルは PS と SS の両強勢音節から学習する。第1段階において文強勢 ST と判定された

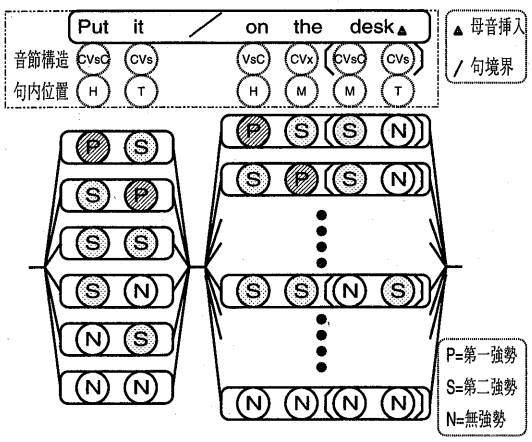


図-2 文強勢誤り候補の生成例

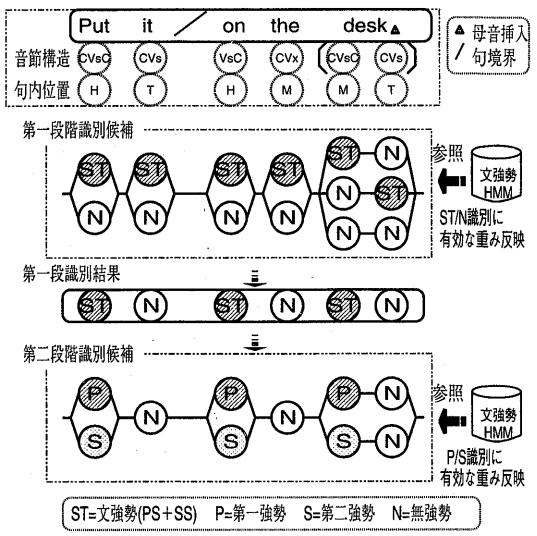


図-3 多段階での文強勢識別

音節に対してのみ，第2段階でPSかSSかの判定を行う。第1段階ではSTとNSの識別に有効な重みを，第2段階ではPSとSSの識別に特化した重みをおのおの反映させることで識別率の向上を目ざす。

3.3.3 判別分析による重みの推定

各段階の識別に有効な重みを推定するために，判別分析を行う。ただし，判別分析ではHMMのようなフレーム単位の特徴量をそのまま用いることが困難なため，音節セグメント単位で代表的な特徴量を抽出した。予備検討[23]において最も判別結果の高かった(1)F_0の音節内平均微係数の絶対値，(2)音節内のパワーの最大値，(3)MFCC 4次元における曖昧母音の分布の中心からの距離（＝母音音質），(4)母音持続長，を特徴量として用いた。ここで各特徴量は$N(0,1)$となるように正規化している。

このように判別分析に用いる特徴量はセグメント単位で抽出し，正規化を行っており，HMMの特徴ベクトルと異なるため，判別分析によって推定された重みをHMMのストリーム重みにそのまま用いることは適格とは言えない。しかし本研究では，すべての重みを同等（＝1）にするベースライン手法よりは，推定重みをおおまかな知覚傾向を示す度合いとして反映する方が望ましいと考え，三つのストリーム重みの総和が3になるように正規化した上で適用することを試みた。

4. 評価実験

4.1 音声試料

表-3に，実験に使用した音声試料に関する情報を示す。米語母語話者の音声は，TIMITデータベースの特定地域（DR 1）の話者31名分を利用した。同一地方の話者に限定したのは，方言による影響を除くためである。ただし，米語における方言の違いは主として一部の母音の相違であり，強勢パターンの差は小さい[24]。本研究においては母音の種類に基づくカテゴリ化に一部影響するが，各方言において十分な学習データがあればモデル化できると考えられる。なお，TIMITの音素バランス（phonetically-compact）文セットの約1/3をカバーしているため音素バランス性は確保されていると考えられる。日本人の音声は，独自に収集した話者8名（男性4名，女性4名）による108文である。

表-3 学習・評価データとラベル

話者	音節	PS	SS	NS	句
母語話者（31名）	3880	572	1045	2263	523
日本人（8名）	935	208	169	558	184

HMMの学習には，3段階の文強勢，音節構造そして句境界に関するラベルが必要である。実際に母語話者が聴覚的に区別できる強勢レベルは高々2,3種類程度であるとの報告[1]もあるため，3段階の識別を必要とする本研究では，専門知識を持つ評価者が望ましい。そこで，米語母語話者・日本人のデータとも，音声学の専門家である[20]の著者に，強勢種と句境界のラベリングを依頼した。ラベリングに際しては，PS, SS, NSの定義（2.3.1項参照）に関する確認を行ったが，作業自体は読上げテキストのみを基に音声データを聴取して行ってもらった。音節構造に関してはTIMITに付属の音素ラベルを利用して決定した。日本人英語の場合は，これから3.2.1項で述べた発音誤り予測を用いて自動補正している。

表-3に示すラベルから日本人と母語話者との間に存在する強勢分布の差異を分析すると，無強勢の割合は同程度にもかかわらず，日本人話者の方が圧倒的にPSの割合が高いことが分かる。これは日本人がピッチアクセントの影響から不必要にPSを増加させていることを示している。

更に母語話者の音声を対象に，音節構造（母音種と構造を別々に分析），及び句内位置が強勢の有無に与える影響について分析した。結果をそれぞれ**図-4, 5, 6**に示す。図-4より，母音種に関しては，曖昧母音（Vx）と無強勢（NS）との相関が非常に高いことが分かる。これは，無強勢の判定に母音の同定が強く寄与していることを示している。実際に米語母語話者の音声では，無強勢母音のうち約52.9%が曖昧母音であった。ただし，曖昧母音のすべてが無強勢にならないのは，音素ラベリングと強勢ラベリングが別の聴取者によって独立に行われていることと，機能語などが曖昧母音のままピッチ変化を受けてPS化する場合があるためである。図-5からは，音節が複雑な構造を持つほど文強勢を受け易いことが分かる。句内位置に関しては，句末の音節が第1強勢となり易いことが示されている。これは特別な強調を含まない文章の強勢パターンでは，文末の強勢が第1強勢になるというノーマルストレスの考えに一致するものである。このように，2.3節で定義したカテゴリが母語話者の強勢分布に対して一定の関係を示し，モデルの詳細化による効果が期待できる。

次に，辞書表記の（孤立単語発声時における）強勢列と比較して，文強勢パターンが受ける変化について分析を行った。結果として，米語母語話者・日本人ともに，辞書表記では複数の強勢を単語内に持つ多音節語でも，文中では最大一つしか強勢音節を持たないことが判明し，3.2.2項の制約が妥当であることが示さ

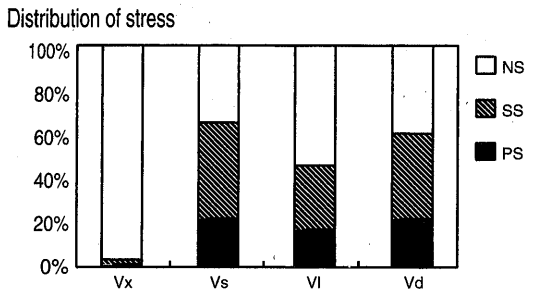


図-4 母音種と強勢の関係

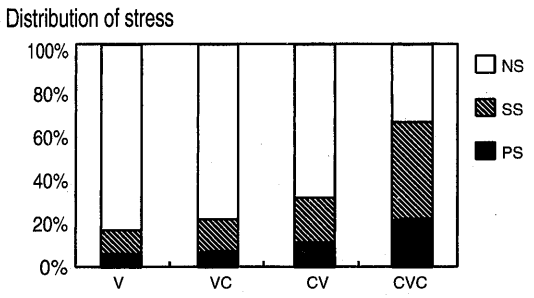


図-5 音節構造と強勢の関係

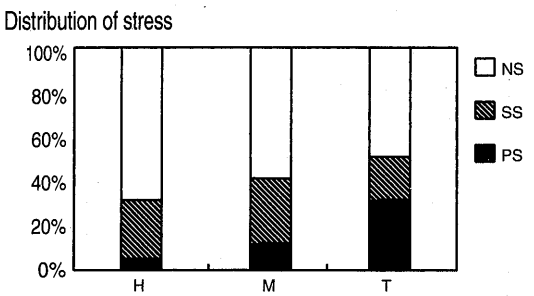


図-6 句内位置と強勢の関係

れた。また，強勢を失った単語の多くは機能語であった。

4.2 文強勢認識実験と考察

本節では，文強勢HMMの学習と認識実験の結果について述べる。米語母語話者の音声を評価する際には，Jack-knife方式によって学習データと評価データを設定した。すなわち31名のうち3名分の音声を評価データとして残りの音声で学習を行う方法を，組み合わせを変えて10通り行った。その評価結果は10回の認識結果を平均した値である。一方，日本人音声を評価する場合は，米語母語話者の全音声を学習用として用いた。初期モデルの学習は，音素ラベルに含まれる時間情報に基づいて音節の始終端を固定して行い，連結学習によって最終的なHMMを構築した。

表-4 種々のカテゴリ化による文強勢識別精度

カテゴリ	モデル数	米語母語話者	日本人
強勢種のみ	3	69.1%	59.3%
音節構造	48	80.0%	65.3%
句内位置	9	77.9%	61.8%
音節構造＋句内位置	144	93.7%	79.3%

ストリーム重みはすべて1.0

表-4に，用意したHMMカテゴリごとの認識結果を示す。ここでは各ストリームの重みはすべて1に固定している。強勢種（3種）のみを考えた場合と，音節構造（16種），句内位置（3種），その組み合わせに応じてPS/SS/NSのHMMを構築した場合を比較している。最も詳細なモデルでは，カテゴリ数の増加に伴って各モデル当たりの学習データが不足する可能性がある。そこでカテゴリを増加するたびに，増加前のモデルを初期モデルとした適応学習を行うことで対応した。

まず，米語母語話者の音声を評価した場合について考察する。最も単純な強勢種のみのモデル化の場合（69.1%）と比較して，音節構造を導入することで10.9%識別率が向上している。これはカテゴリ増加によって，特に強勢に影響を与える音韻現象をより詳細に記述できた結果である。事前に検出する音節構造や母音の種類そのものが，強勢判定に有用な情報を与えていることも考えられる。また，句内位置を導入することで8.8%の識別率向上が見られた。これも音節位置によってピッチ曲線の特性をより詳細に反映できた結果である。更にすべてを組み合わせた詳細なモデルでは93.7%の識別率が得られ，モデル化の妥当性が示された。

次に，日本人話者の音声を評価した場合について考察する。最も単純な強勢種のみのモデルの性能（59.3%）と比較して，音節構造のモデル化によって6.0%の識別率向上が得られた。これは母音挿入等によって音節構造が崩れることの多い日本人英語に対して，音素誤りパターンの導入が効果的に働いた結果と考えられる。一方，句内位置のモデルは2.5%の識別率向上にとどまった。この原因を分析するため，事前の処理として行う句単位の検出率を分析したところ，76か所の句境界のうち，59か所（59/76＝77%）しか適切に検出できていないことが判明した。これは現在の手法が休止情報のみに基づいて句境界検出を行っていることに起因する。そのため今後は句境界の検出率を向上させることが課題である。両方を考慮した詳細モデルでは79.3%の識別率が実現され，日本人音声に対してもある程度強勢識別が可能であることが示さ

表-5 音節構造ごとの識別精度

	V	VC	CV	CVC	計
Vx	85.4	92.7	86.7	86.6	87.7
Vs	92.5	74.7	78.3	69.7	75.2
Vl	88.9	80.2	83.0	77.0	80.4
Vd	87.3	83.1	69.4	76.5	77.0
計	89.3	80.9	79.8	75.9	79.3

表-6 句内位置ごとの識別精度

H	M	T	計
75.2	78.8	86.5	79.3

れた。

この詳細モデルによる日本人音声の識別結果を，各モデルごとに分析した。音節構造に注目した結果を**表-5**，句内位置に注目した結果を**表-6**に示す。音節構造に関しては，特に曖昧母音に対する識別精度が高い。これは4.1節で考察したように，曖昧母音と無強勢の相関が非常に高いためである。ただし表-4において，音節構造のカテゴリ化のみでは最終的な識別精度に大きく及ばないことから，強勢判定全体における母音種の同定の寄与度はそれほど大きなものでない。一方，音節構造が複雑になるほど，識別精度の低下が見られる。現在のHMMは3状態で構築しているが，複雑な音韻変化を持つと考えられるCVC等に対して十分にモデル化できていない可能性が考えられ，状態数を増加させるなどの対応が今後の課題である。次に句内位置に注目すると，句末の音節ほど識別精度が高い。句の最後に存在する強勢はPSとなる場合が多く，文強勢と無強勢で明瞭な音響的差異が生じる傾向があるためである。

4.3 多段階識別の効果

次に，3.3節で述べた多段階識別を導入した。強勢の種類による音響的な差違を判別分析により調べた。米語母語話者の全音声を分析対象として，(1)文強勢(ST)と無強勢(NS)の識別，(2)PSとSSの識別，おのおのに対してどの特徴量の重みが大きいかを分析した結果を**表-7**に示す。比較のため，三つの強勢種を別々の識別対象とした場合の結果も示す。結果から，文強勢か否かの判別にはパワーや母音音質が，PSとSSの判定にはF_0が重要であることが分かる。これは2.3.1項で述べたピッチ変化による影響をPSとSSの違いとして認識する本研究の考えを支持する結果である。

この分析結果に基づいて多段階識別の実験を行った。比較として，1段階識別に判別分析による推定重

表-7 判別分析結果

識別対象	F_0	パワー	母音音質	持続長
ST vs NS	0.170	0.389	0.222	0.218
PS vs SS	0.520	0.175	0.179	0.126
PS vs SS vs NS	0.282	0.306	0.217	0.195

表-8 ストリーム重みの変更による識別精度

識別手法	重み	米語母語話者	日本人
1段階	共通	93.7%	79.3%
1段階	判別分析	93.5%	78.7%
2段階	判別分析	95.1%	84.1%

カテゴリ：音節構造＋句内位置

みを反映させた場合の実験も行った。その場合は，表-7の各強勢種を別々の識別対象とした場合の推定重みを用いた。いずれの場合も推定重みは三つの合計が3となるように正規化している。**表-8**に結果を示す。

まず，1段階識別の結果について考察する。米語母語話者，日本人話者ともに，すべての重みが共通の1の場合と比較して，推定重みを反映させたことによる識別率の向上は見られない。一方，多段階識別を行った場合，米語母語話者において1.4%，日本人話者において4.8%の識別率向上が得られた。強勢種によって知覚に用いる音響的特徴が異なるため，1段階で一律に重みを変更しても知覚傾向を適切に反映できず，2段階識別という枠組みの導入が効果的であることを示している。3.3.3項で述べたとおり，判別分析による重み設定は厳密ではなく，今後の検討の余地があるが，おおまかな傾向でも反映させることの効果が確認できた。

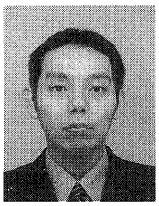
5. ま と め

本研究では，日本人の英語発声を文強勢の観点から評価する方法を提案した。誤りの検出と誤り原因の同定を行うために，日本人の誤り傾向を分類し，(1)強勢段階，(2)音節構造，(3)句内位置に着目したカテゴリ化を行った。HMMによってカテゴリごとの文強勢モデルを構築し，識別実験を行った。日本人音声に対する識別率はベースラインモデルの59.3%に対して，音節構造や句内位置を導入した詳細モデルにより79.3%まで向上し，誤り分類の効果が確認された。更に強勢の種類による音響的な差違に着目し，識別対象に応じて特徴量の重要度を変える方法を提案した。判別分析によって，文強勢と無強勢の識別には母音音質とパワーが，PSとSSの識別にはF_0が重要であることが分かった。この知覚傾向を反映させる2段階識

別によって識別率が更に4.8%向上し，有効性が示された。これらのモデル化は，米語母語話者の強勢検出においても効果的であった。

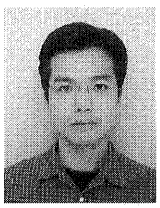
文　献

[1] 渡辺和幸：英語のリズム・イントネーションの指導（大修館書店, 1994），第4-6章.
[2] 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評価. 信学技報, ET 95-44 (1995).
[3] 竹蓋幸生：日本人英語の科学（研究社出版, 1982），第5章.
[4] 竹林　滋, 斎藤弘子：英語音声学入門（大修館書店, 1998），第6, 7章.
[5] 壇辻正剛：IT時代の語学環境としてのCALL. 情報処理, **42**, 1001-1005 (2001).
[6] H. Franco, L. Neumeyer, Y. Kim and O. Ronen: Automatic pronunciation scoring for language instruction. *Proc. ICASSP 97*, Vol. 2, pp. 1471-1474 (1997).
[7] S. Witt and S. Young: Language learning based on non-native speech recognition. *Proc. EUROSPEECH 97*, pp. 633-636 (1997).
[8] C. H. Jo, T. Kawahara, S. Doshita and M. Dantsuji: Japanese pronunciation instruction system using speech recognition methods. 信学論, **E83-D**, 1960-1968 (2000).
[9] 河合　剛, 石田　朗, 広瀬啓吉：2言語の音響モデルを用いた音声認識による非母語発音誤りの検出と発音評価. 音響学会誌, **57**, 569-580 (2001).
[10] F. Ehsani and E. Knodt: Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Lang. Learn. Technol.*, **2**(1), 45-60 (1998).
[11] H. Hamada, S. Miki and R. Nakatsu: Automatic evaluation of English pronunciation based on speech recognition techniques. 信学論, **E76-D**, 352-359 (1993).
[12] S. Hiller, E. Rooney, J. Laver and M. Jack: SPELL: An automated system for computer-aided pronunciation teaching. *Speech Commun.*, **13**, 463-473 (1993).
[13] 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. 信学論, **82-D-II**, 1865-1876 (1999).
[14] A. Waibel: Recognition of lexical stress in a continuous speech understanding system—A pattern recognition approach. *Proc. ICASSP 86*, pp. 2287-2290 (1986).
[15] G. S. Ying, L. H. Jamieson, R. Chen and C. D. Michell: Lexical stress detection on stress-minimal word pairs. *Proc. ICSLP 96*, Vol. 3, pp. 1612-1615 (1996).
[16] G. J. Freij, F. Fallside, C. Hoequist and F. Nolan: Lexical stress estimation and phonological knowledge. *Speech Commun. Lang.*, (4), 1-15 (1990).
[17] K. L. Jenkin and M. S. Scordilis: Development and comparison of three syllable stress classifiers. *Proc. ICSLP 96*, Vol. 2, pp. 733-736 (1996).
[18] 杉藤美代子：日本人の英語（和泉書院, 1996），第1-4章.
[19] P. Ladefoged: *A Course in Phonetics* (Harcourt Brace College Publishers, 1993), Chap. 5, 10.
[20] R. M. Dauer: Stress-timing and syllable-timing reanalyzed. *J. Phonet.*, **11**, 51-62 (1983).
[21] 坪田　康, 河原達也, 壇辻正剛：日本人の誤りパターンの対判別を利用した英語発音教示システム. 信学技報, SP 00-125 (2001).
[22] 堂下修司, 河原達也, 水谷陽一, 児島宏明, 石川雅朗, 北澤茂良：2群対判別法による不特定話者日本語単音節中の子音の識別. 音響学会誌, **45**, 827-836 (1989).
[23] K. Imoto, M. Dantsuji and T. Kawahara: Modelling of the perception of English sentence stress perception for computer-assisted language learning. *Proc. ICSLP 2000*, Vol. 3, pp. 175-178 (2000).
[24] 竹林　滋, 東　信行, 高橋　潔, 高橋作太郎：アメリカ英語概説,（大修館書店, 1988），第3章.



井本　和範

2000年京都大学工学部情報学科卒業。2002年同大学院情報学研究科修士課程修了。同年，㈱東芝に入社。現在，同社研究開発センターにて，音声認識，音声対話の研究開発に従事。日本音響学会会員。



坪田　康

1999年京都大学工学部情報工学科卒業。2001年同大学院情報学研究科修士課程修了。2002年同博士後期課程退学。同年より，京都大学学術情報メディアセンター助手。日本音響学会，情報処理学会各会員。

河原　達也

1987年京都大学工学部情報工学科卒業。1989年同大学院修士課程修了。1990年同博士後期課程退学。同年京都大学工学部助手。1995年同助教授。1998年同大学情報学研究科助教授。現在に至る。この間，1995年から96年まで米国ベル研究所客員研究員。1998年からATR客員研究員。1999年から国立国語研究所非常勤研究員。2001年から科学技術振興事業団さきがけ研究21研究者。音声認識・理解の研究に従事。京大博士（工学）。1997年度日本音響学会粟屋賞受賞。2000年度情報処理学会坂井記念特別賞受賞。情報処理学会連続音声認識コンソーシアム代表。情報処理学会，電子情報通信学会，日本音響学会，人工知能学会，言語処理学会，IEEE各会員。

壇辻　正剛

1979年京都大学文学部卒業（言語学専攻）。1981年同大学院修士課程修了。1984年同博士後期課程単位取得満期退学。同年京都大学文学部研修員。1986年京都大学文学部助手。1990年関西大学文学部助教授。1997年京都大学総合情報メディアセンター教授。現在，同大学学術情報メディアセンター教授。情報学研究科教授兼任。音声分析・CALL（コンピュータ支援型語学教育）の研究に従事。日本音声学会，国際音声学協会，日本音響学会，日本語教育学会，日本言語学会各会員。