音声言語情報処理 29－3
(1999. 12. 20)

英単語音声の強勢音節検出における HMM と DP マッチングの比較

藤澤友紀子　峯松信明　中川聖一

豊橋技術科学大学 情報工学系
〒 441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1
Tel: (0532) 44-6777
E-mail: {fuji, mine, nakagawa}@slp.tutics.tut.ac.jp

あらまし　日本人による英(単)語音声の自動韻律評価を目的として、HMM を用いた強勢音節の自動検出について検討している。本稿では、従来法、及び市販語学学習ソフトにおいて採られている方法論を考慮し、同一単語、異なる 2 発声間における強勢パターンの同一性判定を、HMM、DP マッチング、人間による音響パラメータ視察、の三者間で比較した。判定実験で用いた音声試料には、同一と判定されるべき単語音声対、非同一として判定されるべき単語音声対を同数だけ用意した。判定実験の結果、HMM が最も良い性能を示した。視察、DP では分節的特徴による時間整合や、最適重み付けをしたにも関わらず HMM の性能には及ばなかった。HMM を用いた場合、同一性の判定以外にも強勢位置の情報も自動的に抽出されるなどの利点があり、非母語話者発声に対する韻律評価技術の構築に HMM を用いることの妥当性を示すことができた。

キーワード　HMM、DP マッチング、強勢パターン同一性判定、スペクトル、パワー、ピッチ、継続長

Comparison between HMM and DP matching in terms of their performance of detecting stressed syllables in English words

Yukiko FUJISAWA　Nobuaki MINEMATSU　Seiichi NAKAGAWA
Department of Information and Computer Sciences, Toyohashi Univ. of Tech.
Tenpaku-cho, Toyohashi, Aichi, 441-8580, Japan
Tel: (0532) 44-6777
E-mail: {fuji, mine, nakagawa}@slp.tutics.tut.ac.jp

Abstract　For the purpose of automating the prosodic evaluation of English words spoken by Japanese learners, we have been focusing upon the automatic detection of stressed syllables in English words by using HMMs. In this study, we compared HMMs, DP matching, and human eyes, not ears, in terms of their performance of judging whether two utterances of a single word have the same stress pattern or not. The latter two methods of the judgment were examined because they are conventional methods or have been used in the commercial software for language learning. In the judgment experiment, two kinds of word pair utterances were prepared. One was those which should be judged to have an identical stress pattern and the other was those which should not. And the number of the word pairs were the same between the two. Results showed that the performance was the highest in HMMs though the other two methods utilized additional schemes for the performance improvement. Considering that the judgment by using HMMs give us additional information on the location of the stressed syllables, this study showed the high validity of using HMMs in developing the techniques for prosodically evaluating English words and sentences spoken by non-native speakers.

key words　HMM, DP matching, judgment of stress patterns' identity, spectrum, power, pitch, duration

1 はじめに

非母語話者によって発声された英語音声を評価する場合、「音韻性の適切さ」「単語中の強勢音節の位置及び生成された音響的特徴の適切さ」「文(句)イントネーションの適切さ」といった、複数の異なる言語単位に基づく評価が必要となる。本稿では、評価の対象を英単語に絞っているが、英単語音声の発音を評価する場合、1) 適切な音韻性が生成されていること、2) 適切な位置に強勢が置かれていること、この2点が重要な評価尺度と考えられる。この2つの尺度は音声処理における分節的特徴と韻律的特徴とにほぼ対応付けることができる[1]。

先行研究によれば「不適切な音韻性(r/lの誤りなど)は、それが一貫していれば母語話者は容易に適応できるが、強勢音節位置の誤りはなかなか受け入れ難い」との報告がある[1]。これは、母語話者とのコミュニケーションを考慮すると、韻律的特徴に基付く発音評価を英語の学習者(日本人)に呈示することの重要性を示唆している。また、「日米両言語において単語アクセントの音響的構成/実体が異なるため、日本人は日本語のアクセント生成法、即ち、声の上げ下げに基づいて強勢を生成する傾向がある」との報告例もある[2]。その他にも、ノンネイティブ(第2外国語学習者)の音声認識の研究として、日本人の英語発話において人間が知覚した認識結果とネイティブ発話から学習した不特定話者HMMを用いた認識結果には相関があるとの報告例[3][4]や、継続長を用いると人間の知覚とHMMの相関が高くなるという報告例[5]がある。これらの報告を受け、筆者らは先行研究において、強勢/弱勢に対応する音節単位の音響モデル(HMM)を4つの特徴量(スペクトル、パワー、ピッチ、継続長)を用いて作成し、強勢位置の自動検出を行なった[6]。

こうして現在までに我々は、一貫して強勢音節検出にHMMを用いてきたが、ネイティブと日本人が発声した英単語発声の差を観測・分析することを目的とした場合、DPマッチングを用いた実装も可能である。また、市販されている語学教材ソフトウェアの中には、音響パラメータのみを表示し、判定処理を一切学習者の“目”に任せるものもある。そこで本稿では、異なる2発声間の強勢パターンの同一性判定をタスクとして、HMM、DP、視察の3者間の性能比較を行なった。本タスクでは、同一と判定されるべき単語対と非同一と判定されるべき単語対を同数用意した。HMMにおける判定は、両発声に対して強勢音節検出を行ない、その検出位置を元に判定を行なった。DPの場合は、両発声間のDP距離に対する閾値処理によって同一性を判定した。視察の場合は、ピッチ及びパワーパターンを視覚提示し、それに基づいて同一性の判定を行なわせた。以下各節で、3種類の方法論に基づく判定実験について述べ、その後で性能比較を行なう。

[1] 厳密には、強勢、弱勢間で母音の音質も異なるとされている。

2 HMMにおける同一性判定実験

2.1 強勢/弱勢音節HMM構築に用いた特徴量

文献[7]では、パワー(+Δパワー)、ピッチ(+Δピッチ)、4次のLPC(粗いスペクトル情報)によって、2種類の音節HMM(強/弱勢)を構築し、2音節単語のストレスパターンの認識、識別が行なわれている。また、英単語を発声する場合「強勢のある音節は、強く、高く、長く、発声する」という3原則[8]から、継続時間を考慮する必要がある。

そこで本研究では、4次のメルケプストラム、4次のΔメルケプストラム、パワー、Δパワー、ピッチ、Δピッチの12次元を特徴ベクトルとし、継続時間の制御に基づくHMMを音節を音響単位として作成した。表**1**に、これらの分析条件を示す。パワーとピッチは、対数化した後に、平均値が0になるように正規化している。ピッチの抽出は文献[9]を参考にして行ない、無声部分は、前後の有声区間の終端と始端を直線補間し、その後平滑化してピッチパターンを定義している。また単語語頭/尾に対してはパワーに基づく自動検出を行なっており、HMMとの照合は単語音声区間を対象としている。

2.2 強/弱勢音節モデルの構築

音節単位での強/弱勢HMMは、**2.1**節で述べたように、種々の特徴量を組み合わせてベクトルを構成し、学習した。また、以下に示すように、種々の観点からその音節種類数を制御/定義し、5通りのHMMセットを作成した。なお、HMMは6状態4分布の連続HMMである。

1. 強・弱勢のHMM
   (強弱1種ずつ、合計2種 → 以下 simple と呼ぶ)
2. 対象音節の単語内位置(頭/尾/他)別に構築したHMM
   (強弱3種ずつ、合計6種 → 以下 pos と呼ぶ)
3. 対象音節の構造(単・長母音に分けて、V、CV、VC、CVC構造)別に構築したHMM
   (強弱8種ずつ、合計16種 → 以下 str と呼ぶ)
4. 単語内位置・音節構造別に構築したHMM
   (強弱24種ずつ、合計48種 → 以下 pos-str と呼ぶ)
5. 対象音節の構造(str)、及び、単語内に強勢音節数を1と仮定してa.~c.を考慮して構築したHMM
   a. 対象音節が語頭ならば、後ろの音節の強/弱
   b. 対象音節が語中ならば、前後音節の強/弱
   c. 対象音節が語尾ならば、前の音節の強/弱
   (強(8×3=24)・弱(8×7=56)種、合計80種 → 以下 context と呼ぶ)

表 1. 音響分析条件

サンプリング周波数	12 kHz
量子化ビット数	16 bit
分析窓幅	21.33 [msec](256 samples)
フレーム周期	8 [msec](96 samples)
分析法	14次の LPC 分析

HMM の学習音声試料は、ATR 英単語データベース (話者 em_01、British) 中の 2 音節以上の単語を用いた。単語総数は 3334 であった。

2.3 尤度計算における考察

継続時間長制御に基づくビタビ照合は

$$f(i,t) = \max_{j,\tau}\left[f(j,t-\tau)a_{ji}d_i(\tau)^{\phi}\prod_{k=1}^{\tau}b_i(y_{t+1-k})\right] \quad (1)$$

となる。ここで $f(i,t)$ はビタビスコア (積形)、a_{ji} は状態 j から i への遷移確率、$d_i(\tau)$ は状態 i で τ だけ停滞する確率、$b_i(y_t)$ は状態 i からベクトル y_t を出力する確率密度である。ϕ は $d_i(\tau)$ に対する重みである。ベクトル y_t は **2.1** 節で述べたスペクトル・パワー・ピッチの特徴量を連結して構成されるが、各特徴量間の相関を零と仮定すると、出力確率密度 $b_i(y_t)$ は次のように書ける。

$$b_i(y_t) = \prod_{s=1}^{3}\{b_i^s(y_t^s)\}^{\rho_s} \quad \text{但し、} \quad \sum_{s=1}^{3}\rho_s = 3 \quad (2)$$

ここで、y_t^s は y_t 中、s 番目の特徴量に対応する部分を抽出した部分ベクトル、$b_i^s(y_t^s)$ は y_t^s に対する出力確率密度、ρ_s は各特徴量 (スペクトル、パワー、ピッチ) による出力確率密度に対する重み係数である。**(2)** 式に対して、継続時間制御項を乗ずると次式のようになる。

$$d_i(\tau)^{\phi}\prod_{k=1}^{\tau}b_i(y_{t+1-k}) = d_i(\tau)^{\phi}\prod_{k=1}^{\tau}\prod_{s=1}^{3}\{b_i^s(y_{t+1-k}^s)\}^{\rho_s} \quad (3)$$

結局 **(3)** 式は、スペクトル・パワー・ピッチ・継続長の特徴量が個別に算出する確率 (密度) に対して、各々の重み (ρ_s, ϕ) を乗じてスコアを出していることになる。

ここで、強/弱勢音節モデルが、母語話者 (ネイティブ) 音声を用いて $\rho_s = 1.0, \phi = 1.0$ で既に学習されていたとする。このモデルを用いて、非母語話者による英単語音声中の強勢音節 (話者が意図した強勢位置、辞書的な強勢位置とは必ずしも一致しない) を検出する場合、ρ_s=1.0, ϕ = 1.0 がその検出率を最大にする重みにはならないことは容易に推測される。そこで、この重みを変化させ (即ち、計算機の耳の特性を変化させ)、検出率を最大にする重みを求める (最適重み)。この重みは「強勢生成において利用されている音響的特徴、即ち、学習者の発声上の癖」を反映していると考えられる。筆者らの先行研究 [10][11] では上記考察に基づいた韻律評価手法を実現している。

2.4 同一性判定実験手順と評価用音声試料

同一性判定実験は、日本人話者の発声した単語の強勢パターン (p_j) とネイティブが発声した同一単語の強勢パターン (p_n) の同一性を判定する形で行なわれた。即ち、p_j, p_n 各々に対する強勢音節位置を検出し [6]、それが等しければ同一パターンであり、異なっていれば非同一と判定される。ここで、ネイティブ発声には HMM 学習で用いた学習データを用いた (学習データの検出率は非学習データと比べて高くなるため、ネイティブ発声の強勢パターンは既知であることにほぼ対応する)。本実験を行なうためには、ネイティブによる強勢パターン (辞書的に正しいパターン) と等しいパターンの単語発声及び異なるパターンの単語発声を日本人より収集する必要がある。しかし、発声者が事前知識として持っている強勢位置 (当然誤った知識も含まれる) とは異なる位置に強勢を置いて強制発声させた場合、その発音形態が不自然になる恐れがある。即ち、発声者が持つ事前知識に従って単語発声を行なわせ、その上で、ネイティブによる単語音声と同一、及び非同一の強勢パターンを持つ単語音声の収集が必要となる。

以上の音声試料を収集するため、ネイティブ音声より、以下の 2 点のいずれかを満足する単語を抽出した。

1. 日本人が強勢位置を誤りやすいと考えられる単語
2. 複数の品詞を持ち、品詞に応じて強勢位置が異なる単語

ここで、1. に関しては、文献 [12]～[15] などを参照して選定した。そして、抽出された全単語に対して日本人成人男性 8 人 (評価音声話者) に対して、強勢位置の書き取り試験を行なった。この試験結果を参照することで、各話者の持つ事前知識とネイティブの単語発声における強勢パターンの「ずれ」を知ることができる。そして、各話者毎に、ネイティブ音声と同一強勢パターンを持つ単語群、非同一パターンを持つ単語群を同数定義し、発声させた (発声単語の種類は話者間で同一であるが、各単語に対して付与された強勢パターンは話者によって異なる)。最終的に得られた音声試料は、各話者 65 単語となった。

収録音声は、強勢位置を発声者が決定し、かつ、その通りに意図して発声したデータであるため、強勢の単語内位置情報 (語頭/語中/語尾) を明確に持つと考えられる。**2.2** 節で述べたように、強勢の単語内位置情報を組み入れたモデリングは 2～5 までの 4 種類ある。しかし、3 以降のモデリングは日本人による英語音声からの強勢検出性能を低減させることが示されており [10]、本稿では位置情報のみを反映した HMM (即ち、pos) を使用することとした。

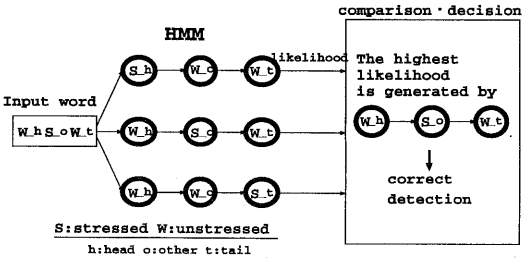

図 1. HMM(pos) の連結法を用いた識別の例

2.5 強勢音節検出方法

検出方法は、先行研究 [6] で提案された手法 (図 **1** 参照) を用いて行なった。図 **1** の S_*、W_* は強勢、弱勢を示している。検出の際には、pos HMM による照合処理が要求する音節情報が与えられる。図 **1** の入力単語は 3 音節であるから、システムは 3 つの HMM 系列を構築する (系列数は、強勢音節を 1 つと仮定しているので入力単語の音節数と等しい)。そして、最高尤度を算出する系列内の強勢位置が強勢検出結果となる。

2.6 実験結果

2.3 節における 4 特徴量に対する重みを全て 1.0 として実験を行なった。前節で述べたように、HMM は pos 形態のものを使用している。本研究は HMM、DP、視察の 3 者間の性能比較を目的としており、かつ、DP、視察に基づく実験では、教師音声と学習者音声間の強勢パターンの同一性を見ることしかできないため、HMM の性能評価も (強勢音節検出率ではなく) 同一性判定率で行なった。同一性判定率は次の式で求めた。

$$\text{同一性判定率} = \frac{\frac{n_s}{N_s} + \frac{n_d}{N_d}}{2} \times 100\ [\%]$$

ここで、n_s, N_s, n_d, N_d は次の通りである。

n_s: 強勢位置が等しいと判定された単語対数
N_s: 強勢位置が等しいと判定されるべき単語対の総数
n_d: 強勢位置が異なると判定された単語対数
N_d: 強勢位置が異なると判定されるべき単語対の総数

同一性判定率を話者別 (8 人、**A**～**H**) に表 **2** に示す. 表より、各話者でばらつきが見られるが、平均で 80%を越える結果が得られている。

3 DP マッチングによる同一性判定実験

3.1 同一性判定実験手順

HMM 使用時と同様のデータベース (同一パターンの単語対、非同一パターンの単語対) を用い、ネイティブと日本人間で DP マッチングに基づく、同一性判定を行なった。ここで特徴量は、HMM 使用時と同様に 4 次のメルケプストラム、4 次のΔメルケプストラム、パワー、Δパワー、ピッチ、Δピッチの 12 次元を用いている。語頭/語尾部分の自動検出は HMM 同様、照合処理の前処理として行なわれており、DP は両端点固定の DP となっている。なお、3 特徴量におけるパターン間距離のレンジが異なり、それが局所距離の計算に影響を与える可能性がある。そこで各特徴量の局所距離は、該当特徴量の分散を用いて (次元毎に) 正規化したものを用いて算出した。分散の値は、ATR 英単語データベース中の約 6800 単語 (2 話者、うち一人は学習話者) を用いて算出した。

DP マッチングにより得られた単語対間のパターン距離に対して閾値処理を行なうことで、同一性の判定を行なった。なお、DP 照合においても HMM 同様、各特徴量におけるパターン間距離に対して重み付けをすることが可能である。但し DP マッチングの場合、継続長重みは定義していない。本実験では、他の三種類の重みを 1.0 とした場合 (第 **2.6** 節に相当) に加えて、話者毎に最適な重みを求めて判定を行なわせた場合の実験も行なった。判定率は第 **2.6** 節で示した手法と同様の手法で算出した。

3.2 実験結果

全重みを 1.0 とした場合を表 **3**、各話者に対して最適な重みを求めた場合を表 **4** に示す。各話者の最適重みの表示は、上からスペクトル、パワー、ピッチ重みとなっている。話者別に重みを最適化することで、平均値 7%ほどの判定率の向上が観測されている。しかし HMM と比べると、重みを操作したにも拘らず、非常に低い結果となった。

表 2. HMM を使用した場合の同一性判定率 [%]

spk.	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	avg.
rate	76.9	83.0	83.5	88.1	68.8	86.2	80.0	89.3	82.0

表 3. DP を使用した場合の同一性判定率 [%]
(話者毎の重みづけなし/
12 次元の特徴量による時間整合)

spk.	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	avg.
rate	62.1	59.3	63.3	56.7	62.6	71.3	50.1	62.0	60.9

表 4. DP を使用した場合の同一性判定率 [%]
(話者毎の重みづけあり/
12 次元の特徴量による時間整合)

spk.	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	avg.
rate	69.3	63.9	67.3	66.7	64.3	77.7	57.9	71.0	67.3
	0.5	0.0	0.5	0.0	0.5	0.0	1.5	0.0	
wgt	1.5	0.5	2.5	3.0	0.5	1.5	1.5	1.5	
	1.0	2.5	0.0	0.0	2.0	1.5	0.0	1.5	

重み (wgt) 表示は上から spectrum/power/pitch 重みである。

表 5. DP を使用した場合の同一性判定率 [%]
(話者毎の重みづけなし/
分節的特徴による時間整合)

spk.	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	avg.
rate	61.9	74.1	64.5	56.1	50.0	88.9	58.4	60.6	64.3

表 6. DP を使用した場合の同一性判定率 [%]
(話者毎の重みづけあり/
分節的特徴による時間整合)

spk.	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	avg.
rate	73.8	75.8	67.3	71.8	66.0	92.2	62.1	65.1	71.8
	0.0	1.5	1.5	0.0	0.0	0.5	0.5	0.0	
wgt	1.5	0.5	1.0	3.0	3.0	2.5	2.5	2.0	
	1.5	1.0	0.5	0.0	0.0	0.0	0.0	1.0	

重み (wgt) 表示は上から spectrum/power/pitch 重みである。

表 **3**、**4** では 1〜10 次元のメルケプストラムを用いないで、直接 12 次元の特徴量に対して DP マッチングを行なった。しかし、ネイティブと日本人では発声における時間構造が大きく異なると思われる。そこで、分節的特徴に基づく時間整合を目的として、1〜10 次元のメルケプストラムを用いて DP マッチングを行ない、最適経路を求め(時間整合)、その最適経路上で 12 次元の特徴量を用いたパターン間距離計算を行なった。この結果を表 **5**(重みは全て 1.0)、表 **6**(最適な重みを使用) に示す。分節的特徴に基づく時間整合を導入し、最適な重みを設定することで約 7%の判定率向上が観測されている。分節的特徴による時間整合の有無で比較した場合、分節的特徴の導入によって判定率は約 5%増加した。

4 視察による同一性判定実験

4.1 同一性判定実験手順

音声情報処理技術を応用した語学学習ソフトウェアは数多く市販されている。しかし中には、音響パラメータの分析・抽出・視覚表示までは行なうが、それに基づいた判定や評価などは一切行なわないものがある。この場合、教師音声 (パターン) と学習者 (パターン) の違いは、視覚表示されたパラメータに基づいて学習者本人が行なうこととなる。このような状況を鑑み、本節では、強勢位置が等しい/異なる単語対に対するパワーパターン及びピッチパターンを視覚的に表示し、両者の強勢パターンの同一性を人間に判定をさせる実験を行なった。なお判定の際に、聴覚提示は行なわなかった。図 **2**、**3** に実験の時に視覚表示したパワーパターン、ピッチパターンの一例を示す (分節的特徴による時間整合あり)。2 発声のパターンが同時に提示されている。学習者はこれらの情報を元に同一性を判定することとなる。HMM、DP 使用時の同一性判定実験と異なり、スペクトル情報は提示しなかった。これは、英会話ソフトでスペクトル情報を視覚提示しているものは (筆者らの知るところ)1 つしかなく[2]、英会話ソフトの模擬という意味では、スペクトル情報は削除した方が望ましいと判断したからである。

実験に使用したデータベースは HMM、DP マッチング時に用いた話者 **A** の 65 単語対であり、判定を行なった被験者は 8 名 (**a**〜**h**) である。この 8 名は評価音声試料収録時の 8 名と一部重複している。特に、被験者 **a** は話者 **A** と同一であるが、聴覚提示がないため、実験終了まで自分の音声 (パターン) であることには気が付かなかった。

4.2 実験結果

同一性判定結果を表 **7**、**8** に示す。表 **7** は、視覚表示時に 12 次元の特徴量に基づく時間整合を行なった結果であり、表 **8** は分節的特徴によって時間整合を行なった結果である。また判定率は、第 **2.6** 節で示した手法と同様の手法で算出している。

[2] The Rosetta Stone/Fairfield Language Technologies Ltd./USA

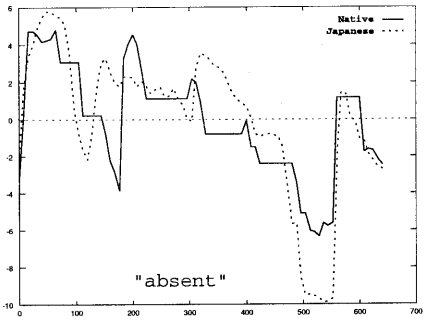

図 2. 視覚表示されたパワーパターン例
(同一強勢パターンの例)

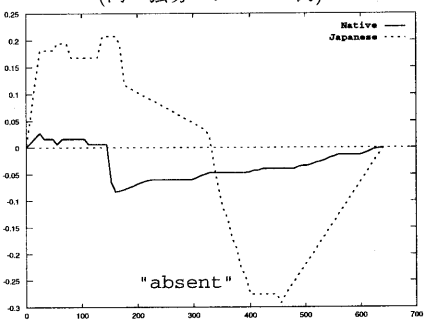

図 3. 視覚表示されたピッチパターン例
(同一強勢パターンの例)

表 7. 視察による同一性判定率
(時間整合なし)

sub.	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	avg.
rate	73.8	79.3	79.8	64.5	79.5	75.7	49.8	74.5	72.1

表 8. 視察による同一性判定率
(分節的特徴による時間整合)

sub.	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	avg.
rate	73.1	79.8	83.6	71.7	77.4	74.0	59.8	89.0	76.1

分節的特徴による時間整合を用いる場合の判定率は DP よりも良い結果となった。視察においても、分節的特徴による時間整合を用いて判定を行なうと良い結果が得られた。スペクトル情報の欠落があるとは言え、この結果はパワーパターンとピッチパターンのみを (時間整合以外に何の処理も加えず) 視覚提示し、学習者に判断を任せた場合でも、学習効果が期待できることを示唆している。しかし、HMM の判定率よりは低い値に止まっている。なお、視察による判定率が DP より上昇した理由としては、次の事が考えられる。DP マッチングの場合、12 次元の特徴量及び、分節的特徴に基づく DP マッチングを行ない

特徴パラメータ間のパターン間距離を計算している。しかし視察の場合は、パワーパターンとピッチパターンの視覚表示そのものは時間整合の結果を反映して表示しているが、2つのパワー(或はピッチ)パターン間の距離の計算は、必ずしも各時刻における局所(パターン)距離を(時間軸方向で)累積しているとは限らず、人間による、より柔軟なパターンマッチング処理が行なわれていることが理由であると考えている。

5 HMM・DP・視察間の性能比較

本節では、HMM、DP、視察による、同一単語/異発声間の強勢パターンの同一性判定実験結果を比較し、得られた知見をまとめる。

- 単語内位置情報のみを導入した強/弱勢 HMM を用いた場合、特徴量重みを全て1.0とした場合でも、判定率は82.0%となった。
- DP マッチングを用いた場合、3特徴量に対して全重みを1.0とした場合、判定率は64.3%となった。
- DP マッチングを用いた場合、3特徴量に対して最適重みを話者別に求めた場合、判定率は71.8%となった。
- 視察による同一性判定では、判定率は76.1%となった。

上記の実験結果より、以下の知見が得られた。

- 各特徴量の重みの適応は、HMM のみならず[10]、DP マッチングでも有効に働く。
- 分節的特徴による時間整合は、DP、視察の両方で有効に働く。
- 各種条件における最高判定率を考えた場合、DP マッチングよりも視察による方が良い結果が得られる。
- しかしながら、HMM の判定率は DP、視察よりも良い値となる。また、HMM による判定処理では、強勢音節の位置情報も付加的に抽出される。

以上の知見より、同一単語異発声間の強勢パターン同一性判定においては HMM を用いた方が DP マッチングや視察よりも性能が良いことを示すことができた。これらの知見は、強勢音節検出処理において、我々が従来より HMM を使用してきたことの妥当性を裏付けるものと考えている。今後、強勢位置検出や強勢パターン比較だけでなく、文(句)イントネーションパターンのモデル化も検討する予定であるが、その場合も HMM によるモデル化は非常に妥当なアプローチであることが示唆される。

6 まとめ

本稿では、同一単語異発声間の強勢パターン同一性判定実験を HMM、DP、視察を用いて行ない、三者の性能比較を行なった。HMM を用いた場合、重み制御などの前処理を行なわない場合において82.0%の判定率を示した。一方、DP マッチングは、最適重みを導入した場合でも71.8%に止まり、視察の場合は、分節的特徴による時間整合を導入した場合でも76.1%の性能しか得られなかった。HMM による処理の場合、教師の見本パターンが必要でなく、強勢音節位置の情報も得られるなど、他の方法では得られない情報も抽出可能であり、HMM の優位性を示すことができた。

参考文献

[1] 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、ET95-44、pp.89-96(1995).

[2] Y. Shibuya, "Differences between native and non-native speakers' realization of stress-related durational patterns in American English," J. Acoust. Soc. Am., Vol. 100, No.4, Pt.2, pp.2725(1996).

[3] J. Bernstein, M. Cohen, H. Murveit, D. Rtischev, M. Weintraub, "Automatic evaluation and training in English pronunciation," Proc. ICSLP'90, pp.1185-1188 (1990).

[4] Y. Kim, H. Franco, L. Neumeyer, "Automatic pronunciation scoring of specific phone segment for language instruction," Proc. EuroSpeach'97, pp.645-648 (1997).

[5] H. Franco, L.Nuemeyer, Y.Kim, O.Ronen, "Automatic pronunciation scoring for language instruction," Proc. ICASSP'97, pp1471-1474 (1997).

[6] N. Minematsu, N. Ohashi, S. Nakagawa, "Automatic detection of accent in English words spoken by Japanese students," Proc. EuroSpeech'97, pp.701-704 (1997).

[7] G. j. Freij, F. Fallside, C. Hoequist, Jr. and F. Nolan, "Lexical stress estimation and phonological knowledge," Computer Speech and Language, 4, pp.1-15 (1990).

[8] 長澤邦紘、「教師のための英語発音」、開文社出版 (1987).

[9] 藤崎博也、広瀬啓吉、瀬戸重宣、"遅れ時間比例窓長の自己相関関数を用いた音声のピッチ自動抽出方式"、信学技報、SP90-86、IT90-102、pp.9-16 (1991)

[10] Y. Fujisawa, N. Minematsu, S. Nakagawa, "Evaluation of Japanese manners of generating word accent of English based on a stressed syllable detection techique," Proc. ICSLP'98, pp.1147-1150 (1998).

[11] 峯松信明、藤澤友紀子、中川聖一、"英単語発音上の癖の自動推定・視覚化とそれに基づく発音能力の韻律的評定"、電子情報通信学会論文誌、Vol.J83-D-II、No.2 (2000).

[12] 上垣暁雄、「即戦ゼミ3大学入試英語頻出問題総演習」、桐原書店 (1992).

[13] 吉田尚志、「高校英語研究 11 月臨時増刊」、研究社出版 (1988).

[14] 文英堂編集部、「大学入試・差をつける英単語・熟語の30時間」、文英堂 (1990).

[15] 文英堂編集部、「大学入試・差をつける英単語と発音の30時間」、文英堂 (1990).