日本語特殊拍の発音自動評価システムとその検討

山本 真人、三輪 譲二

岩手大学　工学部　情報工学科

〒020-8551 岩手県盛岡市上田４-３-５

あらまし： 外国人の日本語音声学習において、習得困難な音声特徴の一つに特殊拍がある。本報告では、日本語母語話者の特殊拍長の知覚実験から得られた特殊拍持続時間正規化関数により、発声速度に依存しない特殊拍長の評価関数を求めた。そして、音声認識技術を用いて学習者の特殊拍の持続時間長を推定し、評価関数を用いてその特殊拍長の良否レベルを求め、学習者にフィードバックするシステムを、Javaアプレットにより作成した。そこで、日本語特殊拍の自動評価システムの構成、及び、本システムを用いた留学生の特殊拍音声の評価結果について考察する。

キーワード： 日本語教育、発音評価システム、特殊拍音声、持続時間正規化関数、Javaアプレット

Computer Assisted Learning System for Japanese Special Mora and Its Evaluation

Masato Yamamoto and Jouji Miwa

masato,miwa@cis.iwate-u.ac.jp

Department of Computer and Information Science,
Faculty of Engineering, Iwate University

4-3-5 Ueda, Morioka-shi, Iwate-ken, 020-8551 Japan

Abstract： This paper describes on a system of computer assisted language learning with Java applet for Japanese special mora speech which is difficult to master for the foreigners. In the system, the duration length of special mora of the learner was automatically estimated using a method of dynamic programming in the speech recognition technology. Then, quality level of the special mora is calculated using a new normalized function for the duration length, which was obtained from perceptual experiment by Japanese natives and is not depend on speaking rate. By the evaluations for 15 foriegn students, both the system and the normalized function are effective for Japanese special mora CALL system.

Key words： Japanese education, Utterance assessment, Special mora speech, Evaluation function of normalized duration, Java applet

1　序論

日本語音声学習において、外国人がより日本語らしい発音を習得するための重要な要素の一つに、特殊拍（長母音、促音、撥音の総称）[1] の習得がある。

特殊拍は、日本語特有の音声特徴 [2] であり、外国人日本語学習者が特殊拍を含む語彙の発音を習得するのは、難しいものとなっている。そのような発音技能を、教育現場で習得することが望まれるが、現実には時間的な制約のために、発音指導に割かれる時間は十分ではない。そこで、教室外でも教師の手を借りずに、単独で発音技能を学べるようにしたいが、学習者自身により発音を評価しながら独習するのは困難である。

以上のような背景をふまえ、音声認識技術を用いて特殊拍の持続時間を推定し、評価関数により特殊拍長の良否レベルを求め、学習者にフィードバックする日本語特殊拍の発音自動評価システム [3] を作成した。

さらに、日本語母語話者の特殊拍長の知覚実験から得られた特殊拍持続時間正規化関数により、発話速度に依存しない特殊拍長の評価関数を求めた。

本論文では、日本語特殊拍の自動評価システムの構成、及び、留学生によるシステムの評価結果について報告する。なお、本論文では、撥音を含まない長母音と促音を、特殊拍と呼ぶ。

2　特殊拍の音響分析

特殊拍と非特殊拍は、語音の持続時間の違いにより区別され、日本語学習者に対する特殊拍の指導には、持続時間の違いを意識させることが重要である。そこで、まず適切な特殊拍長を把握し、学習者の特殊拍長を評価するために、日本語母語話者音声の分析を行い、特殊拍の持続時間分布等を調査した。

2.1　特殊拍のラベリング

単語音声の音響分析を行い、図 1 に示すように、音声波形、パワー、ゼロクロス、スペクトル変化量、ソナグラフと共に、音声パワーとスペクトルの変化量が極大となる時刻を表示する。それらの情報から、視察により音素境界を決定し、音素ラベルデータのファイルを作成した。この音素ラベルは、特殊拍の持続時間の統計量を求めるとともに、4 章で述べる特殊拍学習システムの教師用音素ラベルデータとしても用いた。

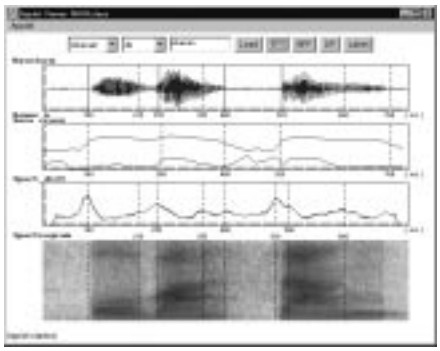

図 1: 音素境界情報の表示例 ([obasan])

2.2　特殊拍の持続時間

日本語母語話者が、特殊拍を含む語彙を発音した時の持続時間について分析するため、男女性日本人アナウンサ 2 人が、非促音と促音（各 10 語）、単母音と長母音（各 14 語）を発音した時の特殊拍 [4] の持続時間の測定を、スペクトル変化量等を参考にして視察により行った。図 2 は非促音と促音、図 3 は単母音と長母音の持続時間を測定した結果であり、表 1、表 2 はその統計値である。この結果、促音および長母音の平均 (標準偏差) は、それぞれ、274ms(24ms)、251ms(37ms) であり、促音より長母音の標準偏差が若干大きいことが分かった。このとき、促音および長母音を含む単語の平均時間長は、それぞれ、722ms、779ms であった。

図 2、図 3 から、非促音と促音の判別のしきい値が約 165ms、単母音と長母音のしきい値が約 170ms であることが分かる。このしきい値を用いると、このアナウンサデータに対しては、特殊拍と非特殊拍を判定できるが、実際の学習者の単語音声の発声時間長は、異なることが多いことから、発声時間長が異なっても判別でき、さらに、その良否レベルをスコアリングすることができる単語長正規化関数が必要である。

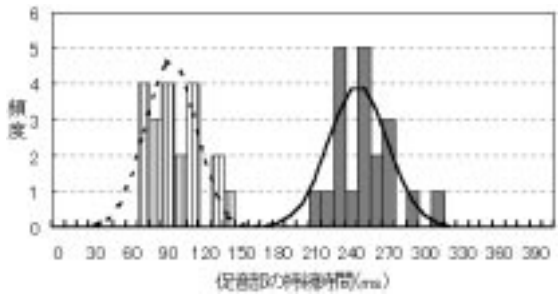

図 2: 非促音、促音の度数分布（アナウンサ）

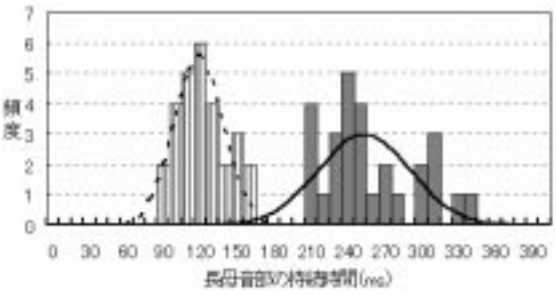

図 3: 単母音、長母音の度数分布（アナウンサ）

表 1: 特殊拍の持続時間の平均と標準偏差（アナウンサ）

	促音	非促音	長母音	単母音
平均値 μ	247ms	91ms	251ms	118ms
標準偏差 σ	24ms	21ms	37ms	20ms
調整倍率 k	1.5	1.5	1.5	1.5

表 2: 単語時間長の平均と標準偏差（アナウンサ）

	促音	長母音
単語時間長 l_w の平均	722ms	779ms

3 特殊拍知覚実験

3.1 知覚実験資料の作成

人間の発声速度は、発話スタイルや個人差によって常に変化するものであり、発声速度に依存しない特殊拍長の評価方法を求めるには、発声速度の変化と特殊拍音声の知覚との関係について調べる必要がある。そこで、促音、長母音の特殊拍の持続時間長を変化させた合成音を用いて、知覚実験を行った。

この実験では、促音の実験単語として、男性アナウンサが発声した「いとう」(3 拍)、「いっとう」(4 拍)、長母音の実験単語として「おじさん」(4 拍)、「おじいさん」(5 拍) の 4 個の刺激単語音声を用いる。これらの音声の発話速度を、0.5〜1.5 倍まで 0.25 刻みの 5 段階 (時間倍率 2.0〜0.67) に変化させた音声を作成する。そして、各音声の語中における特殊拍部分の持続時間を 70ms〜260ms まで 10ms 刻みに 20 段階に伸縮させ合計 400 個の合成音を実験資料として用いる。音声資料の作成には、高品質音声変換プログラム STRAIGHT[5] を使用した。図 4、図 5 に、それぞれ促音、長母音の合成音の例を示す。

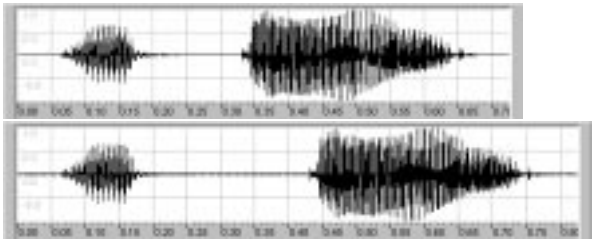

図 4: 「いっとう」の無音部分を伸縮させた音声波形の例（上図：無音 120ms、下図：無音 220ms）

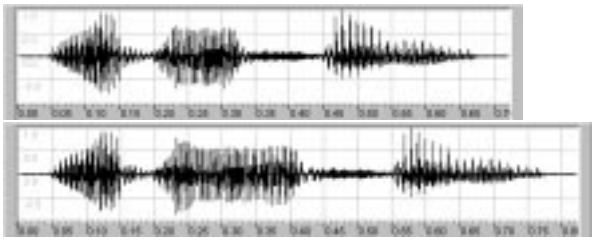

図 5: 「おじいさん」の長母音部分を伸縮させた音声波形の例（上図：長母音 120ms、下図：長母音 220ms）

3.2 知覚実験方法

知覚実験の条件を表 3 に示す。実験では、図 6 に示す Java アプレットで作成した特殊拍知覚実験プログラムにより、特殊拍/非特殊拍の刺激音声がランダムに出力されるので、それを被験者に聞いてもらい、どちらの単語の発音に聞こえたか強制判断してもらった。その判断結果が、ログファイルとして出力される仕組みとなっており、知覚境界を求めることが出来る。

表 3: 知覚実験条件

被験者	日本人母語話者１０人（２０代男性）
実験単語	4 単語 (いとう, いっとう おじさん, おじいさん)
試行回数	２００回／特殊拍組
音声形式	8kHz, 8bits, μ-law 形式

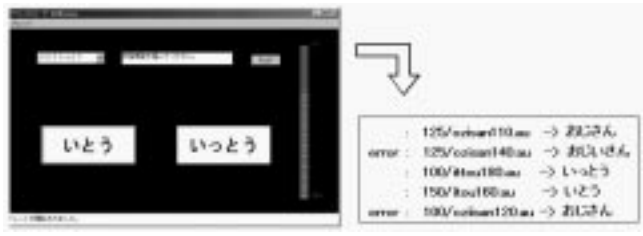

図 6: 知覚実験に使用したプログラムの動作例

3.3 知覚実験結果

図 7 は、「いとう、いっとう」の音声に対し、知覚実験より得られた発声速度ごとの正解率の結果である。

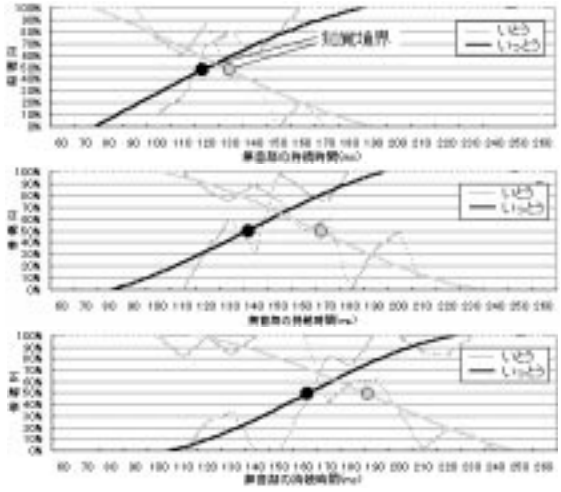

図 7: 促音の単語知覚の正解率（上：時間倍率＝ 0.67、中：時間倍率=1.00、下：時間倍率=1.33）

ここで、図 7 より促音と非促音のそれぞれの正解率が 50%になる持続時間を知覚境界と定義し、発話速度と促音・非促音の知覚境界をグラフで表すと、図 8 のようになる。

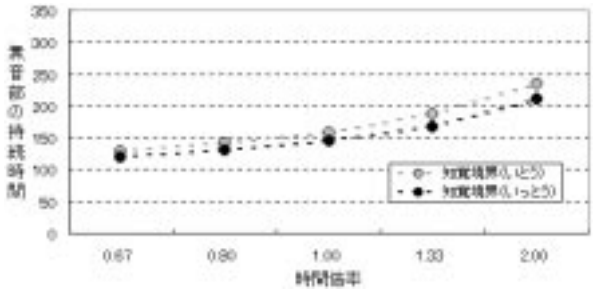

図 8: 促音の 50% 知覚境界と時間倍率の関係

3.4 持続時間正規化関数

図 8 の横軸の時間倍率を単語時間長 (ms) に変換すると図 9 のようになる。ここで、促音の分布が線形である仮定し、回帰直線 ($y = 0.093x + 70$) を求めることにより、単語時間長に対する知覚境界の変化量 0.093 が得られる。

次に、単語時間長に対する知覚境界の変化量と促音長の変化量が同一であると仮定すると、アナウンサの発声より得られた 4 拍の促音の平均単語時間長 722ms

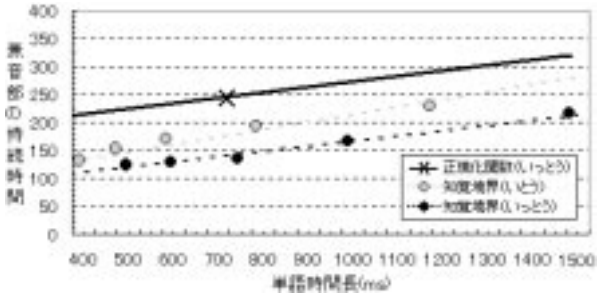

図 9: 促音の知覚境界と単語時間長の関係

と促音平均持続時間 247ms から、促音の持続時間正規化関数を $\mu_Q(l_w)$、単語時間長を $l_w(\mathrm{ms})$ とすると、次式の促音の持続時間正規化関数（4 拍）を得ることができる。

$$\begin{aligned} \mu_Q(l_w) &= 0.093(l_w - 722) + 247 \qquad (1) \\ &= 0.093 l_w + 180 \end{aligned}$$

また、促音と同様の方法により、長母音の単語時間長と知覚境界の関係を表わすと、図 10 のようになる。

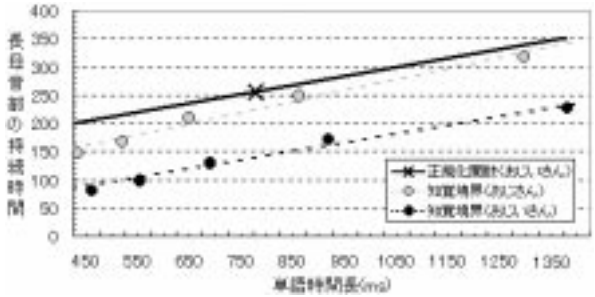

図 10: 長母音の知覚境界と単語時間長の関係

図 10 より長母音の持続時間正規化関数を $\mu_v(l_w)$、単語時間長を $l_w(\mathrm{ms})$ とすると、次式の長母音の持続時間正規化関数（5 拍）が得られる。

$$\begin{aligned} \mu_v(l_w) &= 0.154(l_w - 779) + 251 \qquad (2) \\ &= 0.154 l_w + 111 \end{aligned}$$

式 (1) と式 (2) により、促音と長母音の持続時間を、単語時間長に依存しないように正規化することができる。また、式 (1) と式 (2) の比較より、長母音の方は係数値が大きいので、促音と比較して単語時間長に依存しやすいという結果が得られた。

4 日本語特殊拍の発音評価システム

4.1 システム構成

外国人の日本語学習者に対し、特殊拍の発音習得を支援する自動システムを、汎用性に優れている Java アプレット [7] を用いて作成した。図 11 にシステム構成を示す。本システムは、以下のような特色を持っている。

1. 学習者の発音の良否スコアリング機能
2. 音声認識技術を用いた特殊拍の持続時間の推定
3. 特殊拍音声区間の表示
4. 教師音声の聞き取り
5. マルチプラットホーム対応

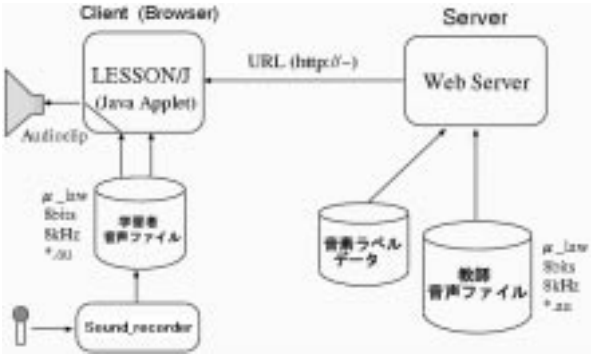

図 11: 特殊拍習得支援システムの構成図

システムの使い方は、まず学習者が自分の学習したい特殊拍を含む単語を選択し、それを発音しサウンドレコーダ等で音声ファイルとして保存する。次に、システム側に置かれている教師音素ラベルデータファイルから、音声認識を用いて対応する学習者の特殊拍の持続時間を推定し、スコア値を表示する。そして、学習者に対して分析結果や矯正フィードバック等の情報が与えられるので、それを基に音長を調整し、適正な発音になるまで練習を繰り返す。

4.2 発音評価方法

本システムでは、学習者の特殊拍の持続時間を推定する際、音声認識技術である動的計画法 (DP:Dynamic Programming)[8] を用いている。DP により、教師の音声と学習者の音声の非線型整合を行った後、バックトラッキングにより最適整合経路を探索することによって、あらかじめ測定してある教師音声の特殊拍の持続時刻ラベルデータから、対応する学習者の特殊拍の持続時間を推定され、学習者の発音に対する評価結果がフィードバックされる仕組みとなっている。

処理手順をまとめると、以下のようになる。

1. 音声データ（8kHz,μ-law）の読み込みと単語音声区間の検出（Load）
2. FFT、16 チャネル帯域化および正規化（FFT）
3. DP による単語整合とバックトラッキング (DP)
4. 特殊拍の持続時間の計測、スコアリング、及び、矯正フィードバック情報の表示（Label）

図 12 に、留学生 A 君による促音の利用例を示す。DP の最適整合経路から、上段で教師、下段で学習者の音声波形と対応する音素区間を表示している。また、上部がコマンドであり、右側がスコア値である。

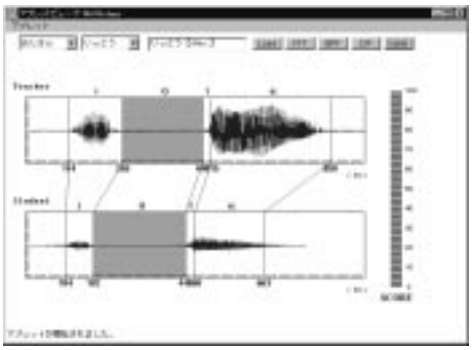

図 12: システムの利用例（促音、留学生A君）

4.3 音声認識技術を用いた特殊拍長の自動推定

本システムにおいて、学習者の特殊拍長は DP により自動推定される。DP により、教師の標準スペクトルパターンと学習者の入力スペクトルパターンの発声速度の時間軸に非線形な伸縮を行い、バックトラッキングによる最適整合経路から、（既知である）教師側の音素境界に対応する学習者の音素境界を推定し、持続時間長の計測を行う。

ここに、その計算手順を示す。格子点 (i,j) には、局所距離 $d(i,j)$ が対応し、以下の式で与えられる。ここで、$\boldsymbol{x}_i$,$\boldsymbol{r}_j$は入力と標準パターン、i,j はフレーム番号、m はチャネル番号を表す。

$$d(i,j) = \|\boldsymbol{x}_i - \boldsymbol{r}_j\| = \sum_{m=1}^{16} |x_{im} - r_{jm}| \qquad (3)$$

DP は、次のような初期条件と漸化式で計算を行う。
初期条件：

$$h(1,1) = 2d(1,1) \tag{4}$$

漸化式：

$$h(i,j) = \min \begin{bmatrix} h(i-1,j-2)+2d(i,j-1)+d(i,j) \\ h(i-1,j-1)+2d(i,j) \\ h(i-2,j-1)+2d(i-1,j)+d(i,j) \end{bmatrix} \tag{5}$$

図 13 は、学習者が“ おじいさん [ozi:san] ”という単語を発音した時の DP マッチングとバックトラッキングの実行例を示している。この例では、母音 i の持続時間を正しく推定しており、学習者の持続時間が短いことが分かる。

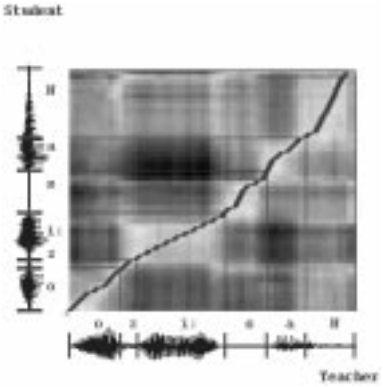

図 13: 最適整合経路の例 [ozi:san]

4.4 特殊拍スコアリング

日本語母語話者の特殊拍長の分析結果を用いて、学習者の特殊拍長の良否を評価する方法として、0 点から 100 点までのスコア値で評価する方法が考えられる。そのようなスコアリングを実現する関数に、事後確率関数に基づく方法 [6] がある。

スコアリング関数を求めるには、まず最初に日本語母語話者の音声分析から得られた持続時間の分布が正規分布と仮定し、平均値と標準偏差から式 (6) と式 (7) により、特殊拍 w_1 と非特殊拍 w_2 の条件付き確率を求める。

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}k\sigma_1} \exp\{-\frac{(x-\mu_1(l_w))^2}{2(k\sigma_1)^2}\} \tag{6}$$

$$p(x|\omega_2) = \frac{1}{\sqrt{2\pi}k\sigma_2} \exp\{-\frac{(x-\mu_2)^2}{2(k\sigma_2)^2}\} \tag{7}$$

ここで、

ω_1	:	特殊拍 (促音、又は、長母音)
ω_2	:	非特殊拍 (非促音、又は、単母音)
$\mu_1(l_w)$	:	特殊拍の持続時間正規化関数
μ_2	:	非特殊拍の持続時間平均値
x	:	推定した音素持続時間
k	:	標準偏差調整倍数

である。

条件付き確率を求めた後、さらに、ベイズの定理に基づいて、先験確率が同一と仮定し、式 (8) と式 (9) より特殊拍と非特殊拍の持続時間の事後確率を求め、これを 100 倍して、特殊拍スコア値 s_1 と非特殊拍スコア値 s_2 を求める。

$$s_1 = \frac{100 \times p(x|\omega_1)}{p(x|\omega_1) + p(x|\omega_2)} \tag{8}$$

$$s_2 = \frac{100 \times p(x|\omega_2)}{p(x|\omega_1) + p(x|\omega_2)} \tag{9}$$

日本語母語話者の分析結果より得られた表 1 の促音の平均値 (μ=247)、促音の標準偏差 (σ=24)、標準偏差調整倍率 (k=1.5) から、スコア関数を求めた結果を図 14 に示す。

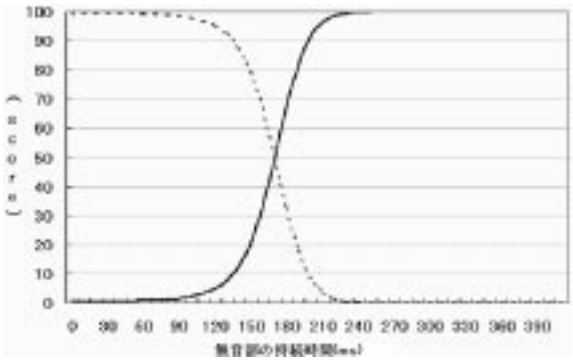

図 14: 促音の事後確率に基づくスコア関数の例

このように、事後確率を求めることにより、発音の良否をスコアリングすることができる。

また、単語時間長と特殊拍長の関係を示す持続時間正規化関数 $\mu(l_w)$ より、単語時間長に応じて特殊拍長の平均値を正規化させることで、発話速度に依存しないスコアリングを行うことが可能になる。

5　システム評価

5.1　留学生音声資料

本システムを評価するために、岩手大学在籍の 15 人の留学生（中国 14 人、バングラディシュ1 人、日本滞在歴約 2 年）に、4 拍の促音 (9 単語)、及び、5 拍の長母音 (3 単語）を含む単語を発音してもらい、合計 180 単語を音声資料とした。

5.2　持続時間とスコア

留学生が特殊拍を発音した時の音声により、持続時間とスコアの関係の考察を行う。

図 12、15 からＡ君の発音を見ると、特殊拍部分を十分な長さで発音していて、スコア値も高く、実際に音声を聞いても正確に発音されていることが分かる。逆に、図 16、17 から、Ｂ君は、特殊拍部分の持続時間の長さが不十分であり、スコア値も低い。Ｂ君の音声を聞いてみると、「おばあさん」が「おばさん」、「いっとう」が「いとう」と発音しているようにも聞こえる。このように、特殊拍の持続時間とスコア値が正しく対応していることが分かる。また、音声波形の表示から、学習者は単語の全体長に照らして相対的な持続時間を知ることができ、教師音声との発音の違いを視覚的に分かりやすく理解することができる。

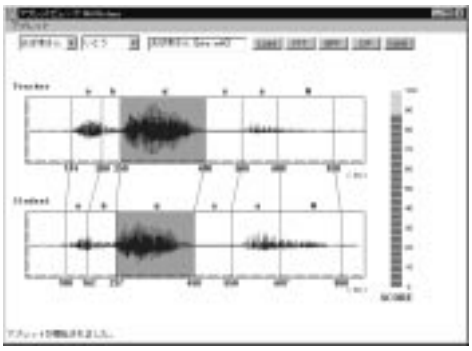

図 15: システムの利用例（長母音、留学生Ａ君）

5.3　持続時間正規化関数の効果

留学生の特殊拍のスコアリングにおいて、式 (6) に持続時間正規化関数を用いた場合のスコア値と用いない場合のスコア値の差の値を縦軸に、単語持続時間を横軸にした時の散布図を、促音は図 18 に、長母音は図 19 に示す。この結果から、単語時間長は約 500ms の開きがあるが、短い単語はスコア値を高くし、長い単語はスコア値を低くしており、単語時間長によりスコア

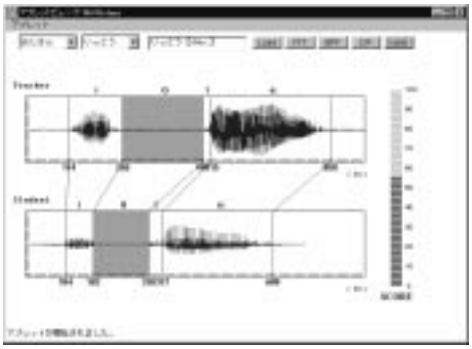

図 16: システムの利用例（促音、留学生Ｂ君）

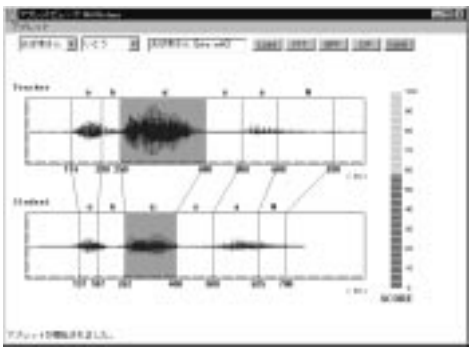

図 17: システムの利用例（長母音、留学生Ｂ君）

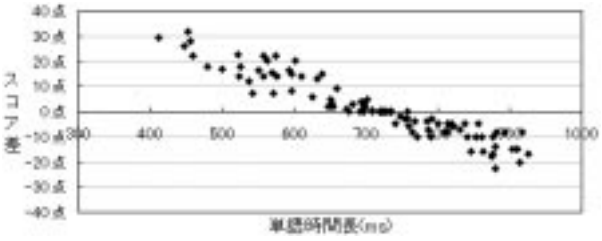

図 18: 促音の単語時間長とスコア差（留学生）

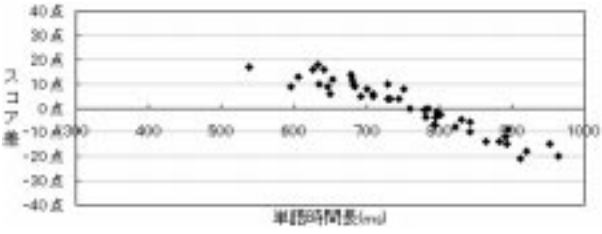

図 19: 長母音の単語時間長とスコア差（留学生）

値を正規化し、単語時間長に依存しないスコアリングをしていることが分かる。

5.4 スコア値の分布

留学生音声に対して、正規化関数を用いたスコアの点数が 90 点未満になったデータは、180 個のうち、促音では全体の約 25%、長母音が約 27%とほぼ同じ割合であった。この 90 点未満のデータにおいて、スコア値の度数分布を、特殊拍と非特殊拍に聞こえる場合に分けて表示した結果を、促音を図 20、長母音を図 21 に示す。

図 20 の促音では、促音に聞こえる場合と促音に聞こえない場合の二つの分布に分離しており、スコア値が特殊拍の学習に役立つことが分かり、また、促音の発音が苦手な留学生がいることが分かる。図 21 の長母音では、長母音に聞こえる場合と長母音に聞こえない場合の二つの分布に分離しており、スコア値が特殊拍の学習に役立つことが分かる。しかし、長母音の誤りは 2 個であり、長母音の発音は得意であることが分かる。

また、スコア値の有効性を検討するため、90 点未満の音声を実際に聞いて判断したところ、促音も長母音も 50 点付近の音声は、特殊拍と非特殊拍のどちらとも判断つかない音声であるが、全体としては、スコア値による評価と知覚による判断がほぼ一致しているという傾向が得られた。

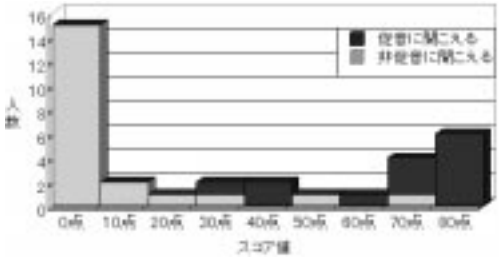

図 20: 正規化関数を用いた促音のスコア値の度数（留学生）

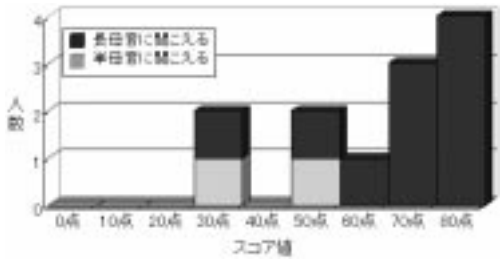

図 21: 正規化関数を用いた長母音のスコア値の度数（留学生）

6 結論

本論文では、外国人の日本語学習者にとって習得が困難な音声特徴である特殊拍（長母音、促音）を取り上げ、発声と知覚の両面から分析を行い、4 拍の促音と 5 拍の長母音の持続時間正規化関数を求め、発話速度に依存しない特殊拍長の評価関数を作成した。

そして、DP マッチングを用いて、既知である教師の特殊拍長と対応する学習者の特殊拍長を推定し、評価関数により学習者の特殊拍長の良否を評価する日本語特殊拍の自動評価システムを作成した。

今回の知覚実験では、4 拍の促音と 5 拍の長母音の発話速度と特殊拍長の関係について分析したが、今後の課題として異なる拍数の促音と長母音の発話速度と特殊拍長の関係についても調査していく必要がある。また、日本語学習期間が短い留学生等に適応し、標準偏差調整倍率 k などを検討する必要がある。

謝辞

音声変換システム STRAIGHT は、和歌山大学河原英紀教授に提供いただいた。また、本研究の一部は、文部省科学研究費補助金・基盤研究 (B)（09558022, 高精度音声分析と音声 CD-ROM を用いた独習用対話型日本語音声教育システムの開発）によった。

参考文献

[1] M. Beckman: "Segment duration and the 'mora' in Japanese", Phonetica, 39, pp.113-135 (1982).

[2] 日本音声学会 : "音声学大辞典", 三修社, 東京 (June 1976).

[3] 山本真人、三輪譲二 : "日本語特殊モーラ長の習得システム", 日本音響学会春季講演論文集, 3-3-3, pp.249-250 (Mar. 1999).

[4] 三輪譲二 : "留学生による日本語音声聞き取り試験とその評価", 日本音響学会 聴覚研究会資料, H99-7, pp.1-8 (Jan. 1999).

[5] 河原英紀、東山恵祐、陸金林、中村哲、鹿野清宏 : "音声分析・変換・合成方法 STRAIGHT の音声符号化への適応について"、信学音声研技報, SP98-10 (1998).

[6] 田嘉鵬、三輪譲二 : "音声教育のための中国語有気/無気音の識別", 電子情報通信学会音声研究会技報, SP97-115, pp.55-62 (Feb. 1998).

[7] Sun Microsystems, Inc.: "Java プログラミング講座", アスキー出版局, 東京 (Oct. 1996).

[8] L. Rabiner and B. Juang : "音声認識の基礎（上）", NTT アドバンステクノロジ (株), 東京 (Nov. 1995).