

## 日本語テ形動詞の発音評価の検討

### On-line Pronunciation Training for Japanese Te-form Verbs

三輪 譲二, Jouji MIWA, 岩手大学・工学部

品川 寛, Satoru SHINAGAWA, ハワイ大学・カピオラニカレッジ

#### 概要

日本語教育において、動詞のテ形は、不規則変化をすることから、文法的に重要な項目である。また、動詞のテ形には、発音が困難な特殊拍の音声が含まれていることから、発音の関門の一つである。このように、動詞のテ形は、文法と発音の両方において、重要な習得項目の一つである。このため、本報告では、自律学習に役立てるため、テ形動詞に含まれる特殊拍音声の自動発音評価とその検討結果について述べる。特殊拍の発音評価実験においては、発声全体の持続時間と拍数などから持続時間を正規化し、事後確率を 100 倍した値を発音評価スコア値として用い、聴覚判断と一致する客観的なスコア値を得ることができた。

**Keywords:** pronunciation training, te-form verbs, on-line evaluation, special mora, duration normalization

#### 1. まえがき

日本語学習者にとって「〜て形」の習得は非常に困難であるが、これを正しく習得し、発音することは、日本語を正しく習得することにおいては、大切な文法パターンである。例を挙げれば、「これを着てください」と「これを切ってください」は、文章で漢字を使って書くと意味は正確に伝わるが、発音のみに頼り、文脈がないと、誤解が生じやすい。このように「〜て形」を正しく習得、及び、発音しないと、意味が正しく伝わらなく、正しいコミュニケーションは成り立たない。この「〜て形」は複雑であり、「〜して、〜いて、〜いで、〜って、〜んで」等、動詞によって複雑に活用が違い、これを正確に習得するのは、日本語学習者にとって非常に困難であり、非常に努力と時間を要する。たとえ、文法的に理解して、それが正しく活用できる（書ける）ようになったとしても、正しく発音されているかは別の問題である。文法上正しく活用できているかは、練習問題をやり、自分で正解と照らし合わせて、正しく解答が出来るかを確認することが出来、自己学習が可能であり、学習者の努力によって、文法的な習得は可能である。

しかしながら、正しく発音が出来るといふことは別次元の問題になってくる。発音は、学習者が一人で練習しても、「正しく発音されているか」「正しく発音されていないか」の判断は自分一人では難しく、学習者が自分で、「自分は正しく発音している」と判断しても、実際には、正しく発音されていないことも多い。これを避けるためには、日本語を母国語とする話者、あるいは、上級の日本語学習者と、ともに練習するか、第三者が実際に発音を聞いて、適切にアドバイスをしてくれる必要性が生じる。このようにつねに第三者に協力を借りて、「〜て形」を正しく発音しているかどうかの判定を依頼することは事実上不可能である。

ここで我々が提案するのが、音声認識エンジンを利用したオンラインの発音評価学習システムである。このシステムを使うことによって、自律学習が可能になる。要するに、学習者は自分が録音した「〜て形」のファイルをサーバーにアップロードし、サーバーが「〜て形」が正しく発音されているかどうかを自動的に判定してくれる学習システムがあると、学習者にとっては、学習者自身のペースで、自律的に、「〜て形」の練習、学習が可能になる。

本報告では、テ形動詞の特殊拍のオンライン発音評価の検討結果について述べる。なお、自動発音評価では、これまで開発した発音評価法<sup>15)</sup>に基づき、音声認識エンジンにより得られた特殊拍の持続時間を特徴量として抽出した後、発声全体の持続時間と拍数などから持続時間を正規化し、事後確率を 100 倍した値を、特殊拍の発音スコア値として算出した。

## 2. テ形動詞音声の発音評価学習システム

### 2.1 システム構成

テ形動詞音声の発音評価学習システムは、文章中の促音などの特殊拍の持続時間からスコアを算出し、その結果を学習者にフィードバックするシステムである。システム構成を図1に示す。図1において、Juliusは、音声認識エンジンであり、音素認識データとその持続時間を得るために用いている。Juliusは、表1に示すように、サンプリング周波数16kHz、フレーム周期10ms、量子化ビット数16bitsとして動作させている。また、発音評価部は、Juliusから得た音素データとその持続時間、発声文テキストを用いて発話速度の正規化を行い、スコアリング処理を行って、図2の例のように、音声波形と認識音素境界を縦棒で表示すると共に、特殊拍の持続時間区間を黒色で表示し、テ形動詞音声の学習を助ける学習者へのフィードバック情報としている。

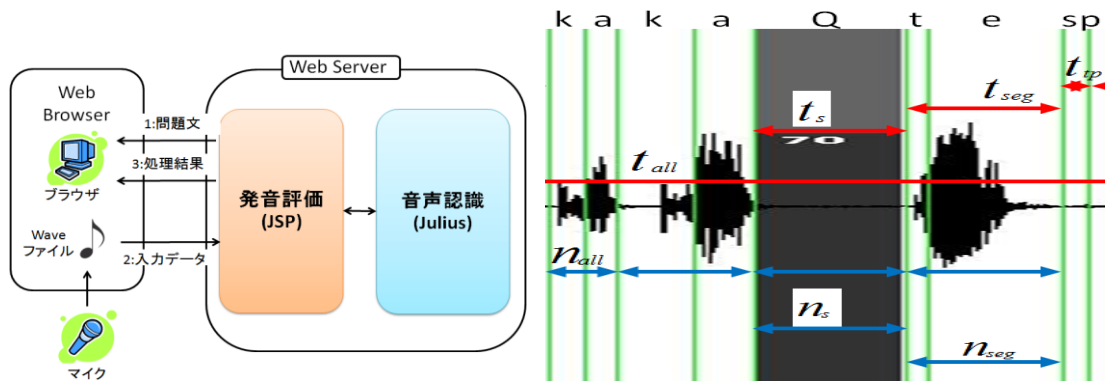


Figure 1 発音評価学習システムの構成

Figure 2 発音評価結果の表示例「かかって」

Table 1 発音評価音声データの形式

ファイル形式 / File Format	WAV
サンプリング周波数 / Sampling Frequency	16 kHz
量子化ビット数 / No. of Bits	Linear 16 bits
フレーム周期 / Frame Period	10 ms

### 2.2 発声速度の正規化

特殊拍の発音評価において、発声速度に依存しない発音評価スコア値を得るため、発声速度の正規化が必要である。なお、発音評価では、音声認識と異なり、真の発音文テキストが予め分かっている。この発声文データは、発音評価時に、評価対象の特殊拍の文中での順番や、総拍数、総文節末拍数、総特殊拍数のパラメータを求めるのに使用することができる。

図2の例において、 $t$ は持続時間、 $n$ は拍数をパラメータごとに分類したものであり、サフックス記号において、 $all$ は全体、 $seg$ は文節末部、 $s$ は特殊拍部、 $p$ は休止部を意味する。このとき、発声全体の平均拍時間  $t_m$  (ms)は以下の式(1)で与えられる。

$$t_m = \frac{t_{all} - \sum t_s - \sum t_{seg} - \sum t_p}{n_{all} - n_s - n_{seg}} \quad (1)$$

すなわち、ここでは、特殊拍や、発声不安定な文節末、発音休止部、特殊拍部を除いて、安定な音声の平均拍時間  $t_m$ を求める。図2の「かかって」の例では、/kaka/という2拍の持続時間の平均時間を計測したことに相当する。

次に、発話速度を正規化した特殊拍の持続時間  $n(x, t_m)$ の算出式(2)に示す。ここで、各変数は、式(1)の平均拍時間  $t_m$ 、日本語の平均的な拍速度  $\mu_t$ 、計測された拍の持続時間  $x$ を表す。本システムでは、 $\mu_t = 121\text{ms}$ として正規化<sup>5)</sup>を行う。

$$n(x, t_m) = x \times \frac{\mu_i}{t_m} \quad (2)$$

### 2.3 スコアリング処理

学習者の特殊拍の良否を評価する際、0点から100点までのスコア値で評価を行い、50点以上が特殊拍音声と判定する。このようなスコアリングを実現するために事後確率に基づいたスコアリング方法を用いた。式(3)にスコアリングの式を示す。

$$S = \frac{100 \times p(\omega_1)p(x | \omega_1)}{p(\omega_1)p(x | \omega_1) + p(\omega_2)p(x | \omega_2)} \quad (3)$$

また、式(3)における条件付き確率は、日本人の特殊拍と非特殊拍の持続時間の分布が正規分布かつ、先験確率が同一であると仮定し、特殊拍の持続時間の平均値を $\mu_1$ 、標準偏差を $\sigma_1$ 、非特殊拍の持続時間の平均値を $\mu_2$ 、標準偏差を $\sigma_2$ として式(3)で定義した。ここで、 $k$ はスコアリング感度係数を示し、 $k=3$ で条件付き確率を求めた。

$$p(x | \omega_i) = \frac{1}{\sqrt{2\pi k\sigma_i}} \exp \left\{ -\frac{(n(x, t_m) - \mu_i)^2}{2(k\sigma_i)^2} \right\} \quad (4)$$

テ形動詞音声の発音評価学習システムの実際的评价結果の例を、図2に示す。このシステムにより評価対象としている拍の現在の持続時間、スコア値が色つきで表示されるため学習者に評価が一目でわかるようになっている。なお、図2の例ではスコア値は70点である。

## 3. 発音評価学習システムの検討

### 3.1 評価対象と評価話者

評価対象データを、表2に示す。本システムに7人の米国人話者の「かかって」と「あつて」という音声の特殊拍のスコアを計測したところ、評価不能だったのが1人、評価できたのが6人であった。評価不能だった理由は、雑音により、認識総拍数が発声文データより多かったためである。

Table 2 発音評価資料データ

話者の国名	米国
最終評価を行った話者数	6名
使用したテ形動詞音声	「あつて/aQte/」、「かかって/kakaQte/」

### 3.2 評価結果

発音評価スコア値は、人間の聴覚判断と一致する客観的なスコア値を得ることができ、正しくスコアリングできていることが確認された。図3に、正規化前と正規化後の持続時間およびスコア値を、話者毎に示す。この図から、正規化後の持続時間と発音評価スコア値の相関が高くなっており、特殊拍の持続時間が、正しく正規化されていることが分かる。

また、「あつて」と「かかって」の話者ごとのスコア値の散布図を、図4に示す。この結果より、スコアに正の相関があり。スコア値の高い話者はどちらの発声のスコア値も高く、スコア値の低い話者はどちらのスコア値も低い傾向があり、複数の評価値を総合することが、発音評価の信頼度を高めることになることが分かる。

### 3.3 検討

テ形動詞発音評価学習では、文法と発音の両方の学習目的があるため、学習意欲を高めることができる。なお、今回は、音声の録音と発音評価がパソコンによる環境で行ったが、今後は、フィードバックの容易な環境での実験が必要である。また、今回は、パソコン付属のマイクを使い、WindowsとMacintoshの両方で利用可能なAudacityを用いて音声を録音した。今後は、図

5 の録音例のように、より利用しやすいスマートホン環境での発音評価学習を行うなどの改善が必要である。さらに、今回は、単独のテ形動詞の発音評価を行ったが、より自然な複数文中のテ形動詞中の特殊拍の発音評価が必要である。

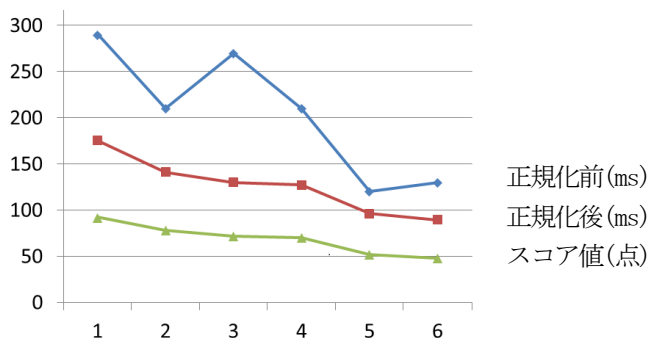


Figure 3 正規化前と後の持続時間及びスコア値の例「かかって」

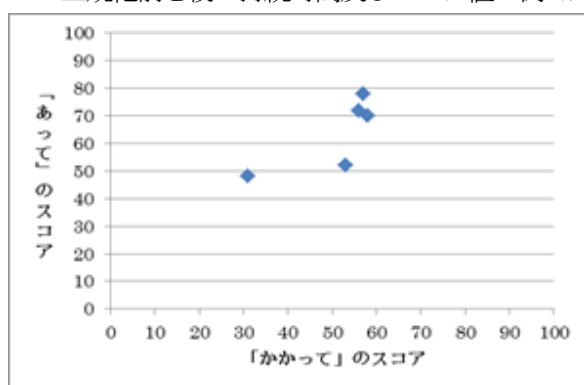


Figure 4 「あつて」と「かかって」のスコア値の散布図の例



Figure 5 iPhone 録音例

#### 4. 結論

テ形動詞音声の発音評価学習の検討を行った。発音評価において、発声全体の持続時間と拍数などから特殊拍の持続時間を正規化し、事後確率を 100 倍した値を発音評価スコア値として用い、聴覚判断と一致する客観的なスコア値を得ることができた。今後は、自律学習として活用し、利用者インターフェースなどをより使いやすくするなどの改善が必要である。

#### 謝辞

本研究を援助いただいた修了生および卒業生の諸氏に感謝します。また、本研究の一部は、科学研究費補助金・基盤研究(B), 22320091, 「デジタル世代の多段自律ネットワーク型日本語学習支援システムの地球規模の運用評価」によった。

#### 参考文献

- 1) 丹野一憲、三輪譲二(2001): インターネット型日本語特殊拍音声の発音自動評価システムとその検討、電子情報通信学会 音声研究会技術報告, SP2000-119, pp.17-24 (Jan. 2001).
- 2) 三輪譲二(2001): 日本語の特殊拍と単語アクセントのインターネット型発音自動評価システムの検討、日本語教育学会春季大会, pp.163-168, 東京女子大学 (May 2001).
- 3) 三輪譲二(2002): インターネット型日本語音声言語教育システムの国際運用実験、日本語教育学会春季大会、お茶の水女子大学 (May 2002).
- 4) 三輪譲二(2004): 発音自動評価機能を有する Web 統合型日本語教育システム、日本語教育国際研究大会予稿集発表 1, pp.36-41 (Aug. 2004).
- 5) 佐藤暢也、三輪譲二(2005): 連続音声中の日本語特殊拍の Web 型発音評価システム、電子情報通信学会 音声研究会, SP2005-138, pp.163-168 (Dec. 2005).