

日本語発話訓練システムにおける自動音声区間検出の有効性

寄元 岳 松尾 光祐 苅木 禎史 宇佐川 毅 (熊本大学)

1 はじめに

近年、様々な分野においてグローバル化が進んでおり、多くの人々が非母語を学習する機会が増加している。話者が非母語でコミュニケーションを行うためには、正確な発話を行うことが重要となってくる。正確な発話を習得するにあたり、学習者は自分の発話に対して客観的な評価を得て、修正を行いながら学習を進める必要がある。そこで本研究では、初級日本語学習者を対象とした日本語発話訓練システムを構築している。本システムは対面式の授業で日本語学習を進めていく上での、補助的な役割を担うシステムであると位置づけている。先に行われた初級日本語学習者の留学生を対象とした本システムの評価実験およびアンケート [1] において、本システムの高い操作性が示された。しかし一方で、音声認識の結果に満足していない被験者もあり、発話訓練の有効性に対する評価が低くなっていた。音声認識の結果が悪くなると、学習者のモチベーションの低下を招いてしまう恐れがあるため、認識精度を改善させる必要がある。

本報告では、被験者の発話音声に対して自動音声区間検出を行うことで、認識精度が改善されることを示し、その有効性について検討する。

2 評価実験

本システムの日本語発話訓練学習では、提示された単語を発話させ正解・不正解をフィードバックする。評価実験では、初級日本語学習者である留学生 8 名を対象としている。8 名の被験者には PC ルームで 50 単語を 3 回ずつ発話させた。発話音声は被験者自身が開始、終了ボタンを押すことで録音される。

2.1 被験者発話音声聴取

実際に評価実験より得られた被験者の発話音声を聴取し、不正解となった原因の切り分けを行った。以下に 3 つの原因を示す。

- (a) 前後の無音区間が長い
- (b) 発話の前半部分が切れている
- (c) 単純な発話ミス

(a) は開始ボタンを押してから発話までの時間、または発話してから終了ボタンを押すまでの時間が長いことに起因し、発話自体は正確に発音できていても、前後の無音区間が長く定常的または突発的な雑音の影響を受けているため、認識結果が不正解となってしまうと考えられる。(b) は PC の性能、および本システムの録音部分に問題があると考えられる。本報告では、(a) の問題点を解決するために自動音声区間検出を行い認識精度の改善を図る。

3 実験

3.1 自動音声区間検出

自動音声区間検出には adintool を使用し [2], 振幅閾値レベルと零交差数を用いて行う。振幅閾値レベルを越える振幅について零交差数をカウントし、それが 1 秒間に指定した数以上になれば、音声区間検出を開始する。検出中に零交差数が指定数以下になれば、音声区間検出を停止する。頑健に切り出しを行うために開始部と停止部の前後にマージンを持たせて切り出しを行う。

3.2 音声認識

本システムの自動音声認識エンジンには、日本語音声認識エンジン Julian を用いている [2]。認識結果は最大 5 候補まで出力する。その中に提示された発音すべき単語が含まれている場合を「正解」とし、含まれていない場合を「不正解」とする。

3.3 実験データ

実験データには前述した評価実験により得られた音声データを用いる。音声データはファイル形式が WAVE audio, サンプリング周波数が 16 kHz, 量子化精度が 16 bit に設定されている。音声データの総数は 1,200 である。

3.4 実験方法

評価実験により得られた音声データを、自動音声区間検出適用前に認識させた場合と、適用後に認識させた場

合とで正解率の比較を行う。正解率は以下のように算出する。

$$\text{正解率} [\%] = \frac{\text{正解数}}{\text{音声データの総数}} \times 100 \quad (1)$$

また、自動音声区間検出のパラメータを表 1 のように設定する。なお、突発的な雑音や発話者の呼気などによって 2 つ以上の区間検出がされてしまった場合、その音声に関しては不正解とする。

4 実験結果

実験結果を表 2 に示す。自動音声区間検出適用前後で正解率が 2.7 ポイント向上している。これより、認識精度の改善が成されたことが分かる。

5 考察

自動音声区間検出の適用により認識精度が改善された理由としては、無音区間が短くなり定常的な雑音による影響が減少したためだと考えられる。図 1、図 2 に区間検出前後の音声波形を示す。両図の横軸は時間、縦軸は振幅を示している。図 1 はある被験者による「はじめまして」という発音の音声波形を示しており、区間検出により不正解から正解となった音声の例である。図 2 は「オーストラリア」という発音の音声波形を示しており、不正解となってしまった音声の例である。これは、発話の後に突発的な雑音が録音されていたために、発話部分と雑音部分の 2 つが区間検出され不正解となってしまった。

6 まとめ

本研究では、被験者の音声認識精度を改善するために、自動音声区間検出を行った。実験の結果、自動音声区間検出を適用することによって正解率が 2.7 ポイント向上し、その有効性を示すことが出来た。今後は、実際に本システムに自動音声区間検出を実装し評価実験を行う。また、雑音環境下における音声の明瞭度改善の手法として多く用いられているスペクトルサブトラクションの導入し、さらなる認識精度の改善を図る。

参考文献

- [1] Junya Watanabe *et al.*, "An Utterance Training System for L2 Learners of Japanese, Implemented on Moodle," Proc. 8th International Conference on Information Technology Based Higher Education and Training, No.91 (CDROM), pp.1-5
- [2] 汎用大語彙音声認識エンジン Julius/Julian, 汎用音声切り出し・録音・転送ツール adintool <http://julius.sourceforge.jp>

表 1: 自動音声区間検出の設定

振幅閾値レベル	500
零交差数	60
マージン [ms]	500

表 2: 自動音声区間検出適用前後の正解率

自動音声区間検出	適用前	適用後
正解数	923	955
正解率 [%]	76.9	79.6

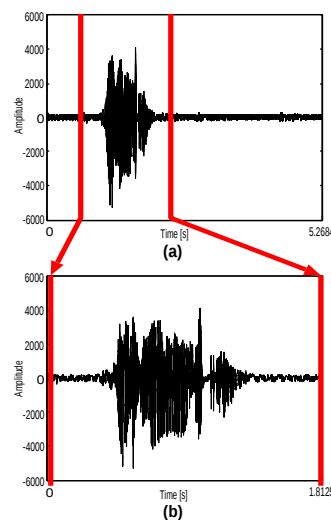


図 1: 正解音声波形例「はじめまして」, (a) 区間検出前, (b) 区間検出後

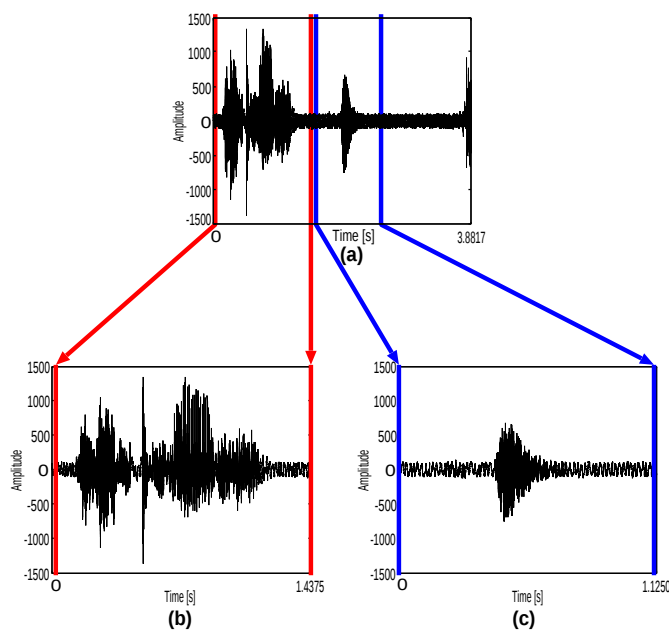


図 2: 不正解音声波形例「オーストラリア」, (a) 区間検出前, (b) 区間検出後の発話部分, (c) 区間検出後の雑音部分