

学習者の音声でフィードバックする英語発音教育システム*

三輪周作, 広瀬啓吉, 峯松信明 (東京大), 山内豊 (東京国際大)

1 はじめに

近年, 日本人の英語発音学習に対する熱意が高まっており, これを急速に普及してきている計算機で援助することができれば非常に有用である. 計算機で発音学習を行う場合, 学習者は自分の発音が正しいかどうかを判断することができないので, 単に英語の教師音声を繰り返し再生するだけでなく, 適切な評価・フィードバックによって学習者を助けることが求められる. このような観点から CALL システムが盛んに研究されている.

われわれはこれまで, 日本語学習者のためのアクセント型発音学習システムとして, 学習者音声の誤ったアクセント型を自動抽出し, 正しいアクセント型に修正して, 学習者へのフィードバックとする CALL システムを提案し, その有効性を示した [1].

本研究では逆に日本人が英語を発音する場合について, 特に問題視されるリズム・アクセントの矯正を行うシステムを開発する. 同様に学習者の音声を修正して提示するシステムとする. システム開発の前の予備的な実験として, 日本人による英語とそれを修正したものに関して英語母語話者による知覚実験を行い, 修正によって英語としての自然さが向上することを確認した.

2 英語と日本語の相違点

日本人が英語発音を不得意とするのは, 日本語と英語の発音体系の違いによる部分が多い [2]. 特にリズム・アクセントについて詳しく見ると, 日本語はモーラ単位の言語であり, 基本的に等間隔に音節 (モーラ) が出現するモーラ等時性を有するとされる. それに対して, 英語はシラブル単位の言語であり, 等時的に強勢が出現する. 強勢間に生ずる弱勢では母音があいまいな音/ə/に近くなり, 子音には脱落や次の音との連結が起こる. これは強勢間に弱勢の音節 (シラブル) がいくつ出現するかにはあまり拠らない. また, 日本語は (撥音以外の) 子音の後には必ず母音が来るが, 英語は子音がいくつも連続することが

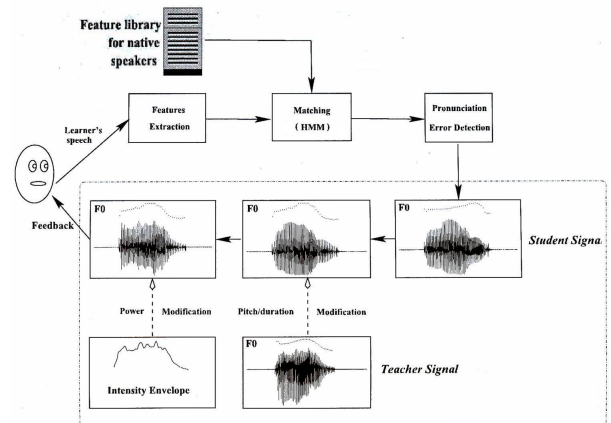


Fig. 1 システムの概念図

許される.

これらの違いが意識できない日本人の英語の発音は, 強勢と弱勢の区別ができておらず, 英語話者にすべての音節が強勢であるかのように聞こえるためにリズムが崩れる. また, 子音が連続するところに母音を挿入するために音節の個数が変わってしまい理解度が大幅に低下してしまう.

3 学習システム

既に提案された, 日本語学習者のためのアクセント型学習システム [1] では, 学習者の間違った日本語アクセント型の音声に対して, 教師の正しい F0 とパワーを TD-PSOLA によってマッピングする. 学習者の声でフィードバックすることのメリットとして, 問題となる点を修正することによってどんな発声になるかを具体的に把握でき, 先生の声真似の練習に陥らない点があげられる.

本研究では, 同様の手法を利用し, 日本人による英語の前記のような問題を直すための CALL システムを構築する. 学習者自身の声で, 母音挿入を修正し強勢弱勢が正しく配置された音声をフィードバックすることで学習者に発音のどこが悪いのかを提示する. 日本人による英語は母音挿入によってリズムが大幅に崩れるため, まず挿入母音の削除を行い, 次に F0・パワーを教師音声からマッピングした. Fig. 1 に構成予定のシステムの概念図を示した. 以下に詳細を示す.

* English CALL System with Feedbacks in Learner's Voice. by MIWA, Shusaku, HIROSE, Keikichi, MINEMATSU, Nobuaki (Univ. of Tokyo) and YAMAUCHI Yutaka (Tokyo International Univ.)

Table 1 音響分析条件

Sampling Freq.	16kHz
Pre-emphasis	0.97
分析窓	Hamming 窓
分析窓長	25ms
窓間隔	10ms
特徴パラメータ	MFCC(12 次) + Δ + Δ Pow(計 25 次)
周波数分析	等メル間隔 Filterbank
Filter Bank	24 チャンネル
CMS (日本語のみ)	発声単位で実行

3.1 発音誤りの検出

リズム・アクセントに着目すると、発音誤りとしては、例えば cat /cæt/ が /cæto/ に変化してしまうような母音挿入によるシラブル構造変化、This is a pen. の機能語である is a をはっきりと発声してしまう強勢位置の不自然な追加などが主に考えられる。それに関連して、本論文ではポーズの除去、挿入母音の除去、F0 の修正、パワーの修正を行った。

3.1.1 母音挿入

当システムでは、HMM による音声認識器を使って挿入母音の検出を行った。日本人の母音挿入誤りには日本語の母音が挿入されることが多く、L1(Native Language) と L2(Target Language) の両方の音響モデルを使うのがよいとする報告 [3] に基づき日米両方の音響モデルを用いた。日本語の音響モデルは文献 [4] 付属の日本語ディクテーション基本ソフトウェアの性別依存 monophone モデル (男女各 2 万文, 132 名) を使用し、英語の音響モデルは DARPA の CSR-I(WSJ0) corpus 男女各 4 千文, 約 50 名) を使って学習したものを用いた。学習に使ったパラメータを Table 1 に示す。

母音を挿入した場合の音素系列が認識時に候補となるように発音文法を作成した。

3.2 フィードバック

学習者の発音の誤った部分を修正するために、同一文を読んだ教師音声を手本として、TD-PSOLA によって F0 の変換、持続時間の変換を行った。あらかじめ HMM による音素セグメンテーションを行い、学習者音声と教師音声の音素

の対応をとり、持続時間については、母音挿入部分は 0 に、他の母音に関しては全体の発話時間に対する比が両者で同じになるように学習者の母音継続長を変換した。F0 については教師の平均 F0 からの変化を、学習者の平均 F0 からの変化にマッピングすることで F0 を決定し、パワーについては教師音声の波形の振幅に合わせて振幅の拡大縮小を行うことで変換した [1]。

4 評価実験

4.1 英語らしさの評価

TIMIT460 文から選んだ 22 文について日本人が読んだものと、母音挿入を削除し、F0 とパワーを教師音声のそれによって修正したものについて米語話者 (CA 出身) 男性 1 名にどちらが英語らしいかを 2 択で選択させた。その際、音声の質については無視して良いと指定した。結果、22 文中 21 文 (95 %) について元の音声より修正後の音声のほうが英語らしいと評価された。

5 まとめ

本論文では、日本人のための英語学習 CALL システムとして、学習者の音声でフィードバックするシステムを提案した。また、挿入音素部分を削除し F0・パワーを教師音声に合わせて変更し、得られた音声に対して、元の音声より英語らしいという評価を得た。ただ、現在の段階では母音持続時間を 0 にした場合の切れ目でのスムージングを行っていない。また、HMM によるセグメンテーションが正確でない部分では音質が極端に悪くなってしまうという問題があり、改善の余地がある。

今後は音声の質の向上を行うためにアルゴリズムの検討を行うとともに、強・弱勢の検出 [5] と修正の機能を加え、スペクトルの変換等も行いたいと考えている。また、実際の学習者に使用して学習者の音声でフィードバックすることに対する有用性の評価を行うことを予定している。

参考文献

- [1] Gendrin *et al.*, 音講論 (春), 371-372, 2003
- [2] 竹林滋, “英語音声学” 研究社, 1996
- [3] Kawai Ph.D. dissertation, Univ. of Tokyo
- [4] 鹿野 他, “音声認識システム” 情報処理学会, 2006
- [5] 小橋川哲 他, 信学技報, SP2001-100, 2001