

英語発話学習に向けた音素連結・脱落・同化パターンの分析と検出

情報科学科 平田 里佳

指導教員：入部 百合絵

1 はじめに

英語の発音学習のための教材は CD や書籍、最近では外国語発音の自学自習を目的とした CALL (Computer Assisted Language Learning) 教材など様々なものが存在する。しかし、これらの教材では孤立単語の発音学習が中心であるため、文章のような複数の単語を発音する際に生じるリンキング (単語連結時に生じる音素の連結, 脱落, 同化) の発音学習には対応していないことが多い。リンキングとは、例えば did you のように二つの単語を発音する場合に、did の最後の音素 /d/ と you の最初の音素 /y/ が連結することで異なる音韻となるような現象を指す。リンキングは英語らしい発音や英語の聞き取りにおいて重要である。従って、本研究では単語連結時に生じる連結・脱落・同化に対する英語母語話者および日本語母語話者のパターンを分析し、リンキング箇所に対する音声認識を試みる。これにより、誤った発音を適切に指摘できるようにする。

2 リンキングパターンと日本人の発音誤りパターン

2.1 英語母語話者のリンキングパターン

リンキングはその様式により、次の 3 つに分類されている [1][2]。

- **連結 (Liaison)**
単語の切れ目が変化する (例: put on [put an] → [putan])
- **脱落 (Elision)**
破裂の省略や子音の脱落が起こる (例: could be [kəd bi:] → [kə^dbi:])
- **同化 (Assimilation)**
音素の変化や音の挿入が起こる (例: did you [did ju:] → [diɔ̃ju], may I [mei ai] → [meiⁱai])

2.2 日本語母語話者の発音誤りパターン

日本語母語話者による英語の発音誤りは、置換誤りと母音挿入誤りの大きく分けて 2 種類存在する [3]。

- **置換誤り**
発音が類似した音素同士を置き換える (例: /ɑ/, /ʌ/, /ə/ と日本語「ア」の混同, /l/ と /r/, /θ/ と /s/ の混同)
- **母音挿入誤り**
子音の後ろに不要な母音を挿入する (例: it [it] → [itoJ*])
*日本語の「オ」

3 リンキングパターンを考慮したネットワーク文法の作成

本研究で使用した音声認識システムでは、はじめに入力音声から MFCC, ΔMFCC, ΔΔMFCC, パワー, Δパワー, ΔΔパワーの計 39 次元の特徴量を抽出する。

次に、抽出した特徴量から学習した音響モデルを構築する。このとき、HMM (Hidden Markov Model) から得られる音素列に対して、音響モデルと言語モデルを用いて尤もらしい音素を選択する。一般的な発音学習における音声認識器では、音響モデルに英語母語話者による英語音声のみを学習させるが、この場合日本語特有の音素を認識できない。本研究では、音響モデルに英語母語話者による英語音声 (TIMIT) に加え、日本語母語話者による日本語音声 (CSJ) も学習させることで、日本語特有の音素にも対応できるようにした。言語モデルには、本研究で明らかにしたリンキングパターンを組み込んだネットワーク文法を用いた。本研究では、図 1 のように正解音素 (白色) と誤り音素 (灰色) をネットワーク状につないだ文法を作成した。以上のように、認識候補を限定し、リンキングパターンに沿った音声認識を目指す。

4 評価実験

ネットワーク文法の使用による認識精度の向上を明らかにするため、言語モデルにネットワーク文法を使用する場合としない場合での認識精度の比較実験を行った。

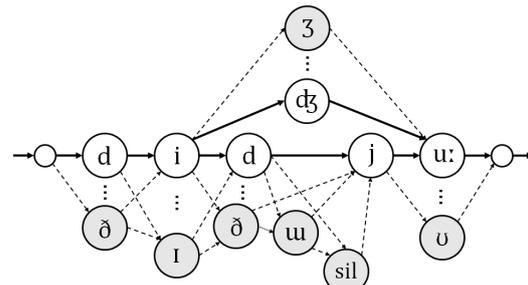


図 1 誤り候補を含んだネットワーク文法の例「... did you ...」

表 1 ネットワーク文法の有無に対する認識結果

| ネットワーク文法 | 認識率 (%) | 認識精度 (%) |
|----------|---------|----------|
| なし | 19.75 | 12.90 |
| あり | 41.40 | 39.65 |

4.1 構築した音響モデル

使用した音声試料は以下の通りである。

- **学習データ**
 - TIMIT: 男性 438 名, 4,380 文
 - CSJ: 男性 24 名, 6,231 発話
- **評価データ**
日本人による英文読み上げ音声 UME-ERJ: 男性 48 名
90 単語列

本研究で構築した音響モデルは 5 状態 3 ループの left-to-right 型である。また、16 混合 monophone とした。

4.2 実験結果

ネットワーク文法の有無に対する認識結果を表 1 に示す。結果から、ネットワーク文法を使用することで認識率および認識精度が向上した。特に認識精度が大幅に向上したことから、認識時の挿入誤りを大幅に削減できたといえる。

しかし、ネットワーク文法を使用した場合でも認識率、認識精度共に 40% ほどであり、十分に認識ができていたとはいえない。正解率の内訳を確認すると置換誤りが圧倒的に多いことがわかった。特に、英語母音を日本語母音に置き換えてしまう場合や、/t/, /p/, /b/ に対する置換誤りが多く存在した。特定の音素に置き換わってしまう問題に対しては、ネットワーク文法のパスに重みを設けるなどして、認識する音素列に偏りが生じないように改良する予定である。今後はネットワーク文法の見直しや音響モデルの改善などを行い、認識率および認識精度の向上を目指す。

5 おわりに

本研究では、従来の発音教材や音声認識利用の英語発音教示システムでは重要視されていないリンキング (単語連結時に生じる音素の連結, 脱落, 同化) に対する英語母語話者および日本語母語話者のパターンを分析し、リンキング箇所に対する分析と音声認識を試みた。音声認識器を構築する際に、音響モデルに英語母語話者による英語音声に加え、日本語母語話者による日本語音声を学習させ、さらに言語モデルとして誤り候補を含んだネットワーク文法を組み込むことで、認識率および認識精度が向上した。しかし、ネットワーク文法を使用した場合でも、十分な認識結果とはいえず、改良が必要である。

参考文献

- [1] 巽一朗, “英語の発音がよくなる本”, KADOKAWA, 2005 年.
- [2] 深澤俊昭, “英語の発音パーフェクト学習辞典”, アルク, 2000 年.
- [3] 坪田康他, “日本人の誤りパターンの対判別を利用した英語発音教示システム”, 信学技報, SP2000-125, 2001 年