

## 日本人の誤りパターンを考慮した英語発話中の音素連結・脱落・同化検出

野村 知里<sup>1</sup> 入部 百合絵<sup>2</sup> 桂田 浩一<sup>3</sup> 新田 恒雄<sup>3</sup>豊橋技術科学大学 知識情報工学課程<sup>1</sup>  
豊橋技術科学大学 情報メディア基盤センター<sup>2</sup>  
豊橋技術科学大学 大学院工学研究科<sup>3</sup>

## 1. はじめに

言語にとって音声は不可欠であるが、日本の語学教育は読むことと聴くことに偏重しており、外国語の習得に必要な「発音」の教育を軽視する傾向にある。特に、言語特有の発音を習得するために必要なリズム、イントネーション、リンキング（単語連結時に生じる音素の連結、脱落、同化）の学習時間が絶対的に不足している。特に、英語の発音に関しては、リンキングが重要であるが、近年開発されている音声認識技術を用いた発音学習システム[1]においても、音韻や孤立単語の学習支援が主たるものであり、リンキングの学習支援機能はほとんど存在しない。

本研究では英語発話中に生じるリンキングのパターンを調査し、そのパターンを組み込んだ音素認識器を開発する。特に、本研究では日本人のための英語発音学習を対象とするため、日本人英語学習者のリンキングに対する誤り傾向を含んだネットワーク文法を構築し、音素認識器に組み込む。これにより、日本語音素混じりの日本人英語音声に対しても、高精度に音素間の連結・脱落・同化を検出し、その発音誤りを的確に指摘することを目標とする。

## 2. 英語発話中のリンキング

英語母語話者は単語毎に区切ることなく連結して、あたかも一つの単語のように発音することが一般的である。このように、単語間の音素が連結され、発音に何らかの変化が生じる現象をリンキングという[2][3]。

## 2.1 英語母語話者のリンキング規則

英語母語話者が発話する際に生じるリンキングの様式として以下の3つに分類される。

- (1) 連結(Liaison)：単語の切れ目が変化する  
例： can I , look it up
- (2) 脱落(Elision)：破裂の省略や子音の脱落が起こる  
例： take care , could be
- (3) 同化(Assimilation)：音素変化や音挿入などが起こる  
例： did you , thank you , may I

これらの様式の種類は、単語連結箇所の調音様式によって決定される。例に示したように連結は連結箇所の前後の音素が子音と母音の際に起きる。同様に、脱落は前後の音素が子音と子音の場合に生じ、同化は子音と半母音、母音と母音の組合せの場合に起きる傾向にある。

Detection of Liaison, Elision, and Assimilation in English based on Japanese Pronunciation Patterns.

Chisato Nomura<sup>1</sup>, Yurie Iribe<sup>2</sup>, Kouichi Katsurada<sup>3</sup>  
and Tsuneo Nitta<sup>3</sup>

Department of Knowledge-based Information Engineering, Toyohashi Univ. of Tech.<sup>1</sup>

Information and Media Center, Toyohashi Univ. of Tech.<sup>2</sup>

Graduate School of Engineering, Toyohashi Univ. of Tech.<sup>3</sup>

## 2.2 日本語母語話者のリンキング誤りパターン

日本人は単語間を区切って発音する傾向が強いが、英語らしい発音を目指すためにはリンキングの習得が必要不可欠である。本研究は、リンキング部分を中心とした日本人の英語発音の誤り傾向を考慮した音声認識器を開発するため、その誤りパターンを調査した。以下に誤りパターンをまとめる。

- 単語間を連結しない
- 日本語に存在しない音素を日本語の音素に置き換えて発音する
- 子音で終わる単語の後に日本語の母音を付けて発音する [4][5]

本研究では以上の傾向を踏まえた音素列をネットワーク文法上に構成することで音素認識候補を限定し、認識率の向上を図る。

## 3. 音素認識器の概要

図1に調音特徴を用いた HMM (Hidden Markov Model) に基づく音素認識器の概要を示す。入力された音声は特徴抽出部において調音特徴(AF)に変換される。ここで調音特徴とは単音分類に用いられる調音位置（舌の位置や口唇の高低）と調音様式（発声に要する舌、口唇、口蓋などの調音器官の動作、例：摩擦音、破裂音など）の諸属性を指す。本研究は日本語音素混じりの英語音声を対象としているため、日本語と英語の音声データから成る2種類の多層ニューラルネットワーク(MLN)を構築する。そして、入力音声の各音素が日本語音素あるいは英語音素のどちらに近いのかを判別するために、多層ニューラルネットワークから出力された2種類の調音特徴の値(0~1)と、各ニューラルネットワークの教師信号との距離を求め、距離の小さい調音特徴を選択しHMMに通す。

日本人に多い単語間の連結しない誤りを正確に指摘するためには、連結の有無を精度よく検出する必要がある。それを判断する材料の一つとして、単語間に生じる無音区間の検出が考えられる。本研究では無音を抽出するために、前後音素に応じて無音の種類を分類し（例えば、有声/無声破裂音の前に生じる無音は vcl, ucl, 摩擦音と半母音または鼻音の間に生じる無音は epi とした）、それらを含む日英 63 音素を音響モデルで学習した。また、リンキングが生じている音素列候補と生じていない音素列候補を別に設けることで誤り候補を組み込んだネットワーク文法を構築する。ネットワーク文法とは、図2のように正解音素列（白色）と誤り候補（灰色）をネットワーク状につなげたものである。図2は IPA（国際音声記号）を用いて表現したものである。ネットワーク文法では音響尤度を計算することで、設定した音素列候補の中から適切な音素列を選択する。最終的に音響モデルとネットワーク文法から構成された音素認識器に通すことで認識結果を得る。

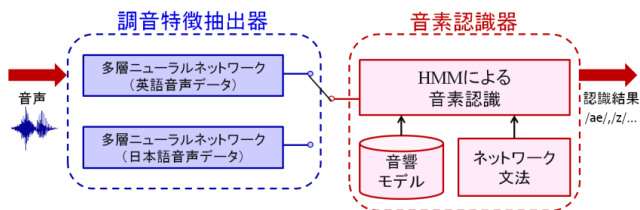


図1. システム概要図

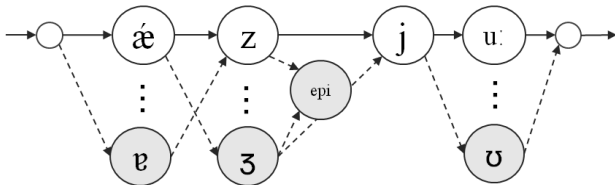


図2. 誤り候補を含んだネットワーク文法の例「...as you ...」

#### 4. 評価実験

本研究で開発した日本語人の英語発音誤りパターンを考慮した音素認識システムの評価実験を行った。実験ではリンキング規則と誤りパターンを含むネットワーク文法の効果を検証するため、それらの有無による比較実験を行った。

##### 4.1 音声試料

実験で用いた音声試料は以下の通りである。

- ・学習データセット  
英語読み上げ音声 TIMIT2608 文 男声 326名  
日本語新聞記事読み上げ音声 JNAS 4962 文 男声 33名
- ・評価データセット  
日本人による読み上げ英語音声データベース ERJ 92 単語列 男声 48名[6]

##### 4.2 実験結果

図3にネットワーク文法による音素列制約有無の実験結果を示す。日本人英語音声に対しても、文法を用いたことで正解率、正解精度共に向上した。特に正解精度が向上したが、これは認識候補を制限することで、湧き出し誤りが減少したためだと考えられる。

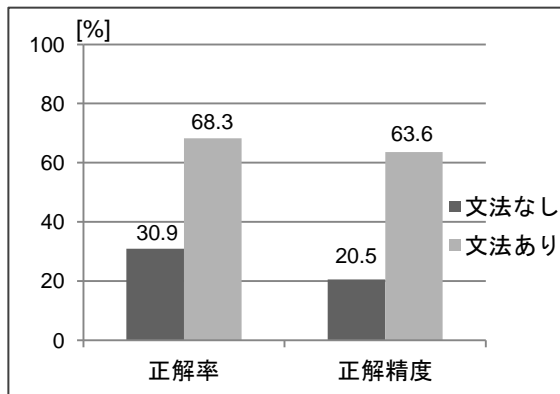


図3. ネットワーク文法の有無による音素認識性能の比較

表1. 日英MLNから抽出した調音特徴の選択率 [%]

テストデータ	日本語 MLN	英語 MLN
JNAS	14.07	85.93
TIMIT	27.99	72.01

しかし、ネットワーク文法の尤度計算において話速が速いものは、リンキングの音素列候補の方がリンキングの生じていない音素列候補よりも高くなる傾向があった。そのため、話者がリンキングによる変化をしていなかったとしても、リンキングによって発生する脱落があったと判定されるといった誤認識が生じた。今後は話者による話速の違いを考慮していく必要がある。また単語末の有声破裂音 (/b/, /d/, /g/) が無声破裂音 (/p/, /t/, /k/) として誤認識される傾向も確認された。単語末は呼気が弱くなることが多いためである。

今回の結果は、正解率がネットワーク文法無の場合30%、有の場合でも68%と全体的に低い。その原因の1つと考えられる2種類のMLNから得た調音特徴の選択手法に関する検証を行った。JNAS, TIMITのテストデータに対し、日英ニューラルネットワークのどちらの出力結果(調音特徴)が選択されたかを表1に示す。これより、テストデータTIMIT(ネイティブによる英語音声データ)に関しては、英語音声によって学習されたMLNから出力された調音特徴を選択していることが分かる。従って、TIMITに対する調音特徴の選択は正しいと判断できる。一方JNASは日本語音声データであるのにも関わらず、英語MLNから出力される調音特徴を多く選択している。つまり、日本語音素を英語音素と判断している可能性が高い。従って、日本人話者の英語音声データERJのように英語音素と日本語音素が混在する場合、日本語固有の音素を正しく選択できていないことが予想される。それ故に、日本語音声データに対する調音特徴抽出精度を向上させることが必要である。

#### 5. まとめ

英語母語話者のリンキング様式および日本語母語話者の英語誤りパターンの調査を行った。調査結果に基づいたネットワーク文法を構築し音素認識器に組み込むことで音素認識性能が約40%向上した。今後は調音特徴の抽出手法及び選択方法を検討し、非母国語話者の音声に対する音声認識性能の向上を目指す。

#### 参考文献

- [1] 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)
- [2] 巽一朗, “英語の発音がよくわかる本” (中経出版, 2005)
- [3] A.Stephen Gibbs, “The Liaison of English Part One”, 関西大学外国語教育研究 16, 27-46 (2008)
- [4] 野田哲雄, “英語発音誤りと変異音”, 東京学芸大学紀要 2 部門 36, 97-112 (1985)
- [5] 竹林滋, “新装版 英語音声学入門” (大修館書店, 2008)
- [6] 峯松信明, 他, “日本人話者による英語文・単語音声データベースの構築”, 日本音響学会秋期講演論文集, 1-Q-30, 199-200 (2001)