

非母語音声の認識のための実音声を用いた発音辞書獲得

辻岡 聡[†] リアン ルー^{††} サクリアニ サクティ[†] グラム ニュービグ[†] 戸田 智基[†]
中村 哲[†]

[†] 奈良先端科学技術大学院大学 〒630-0192 奈良県生駒市高山町 8916-5

^{††} エディンバラ大学, 音声技術中央研究所, Edinburgh, EH8 9AB, UK

E-mail: †{tsujioka.satosi.tl4,ssakti,neubig,tomoki,s-nakamura}@is.naist.jp, ††liang.lu@ed.ac.jp

あらまし 国際会議などでは英語が国際標準語として用いられ、英語非母語話者の間でも英語で意思疎通を図る場面が多い。このような非母語音声認識して議事録を作成するなどの応用技術を考えて場合、非母語音声認識を高精度に行う必要がある。しかし、非母語話者の音声は母語話者に比べ、発音の揺らぎやブレが原因となり、非母語音声の認識精度は母語音声よりも低下する問題が生じる。非母語音声認識において、音響モデル、発音辞書、言語モデル、デコーディングの各処理系にて考慮する必要があるが、本研究では発音辞書に焦点を当てる。非母語話者の発音の揺らぎに対処するために、単語表記列から発音系列候補を予測する G2P (Grapheme-to-phoneme) ツールを用いて、複数の発音バリエーションを生成し、非母語音声話者の実音声から生起頻度の高い発音バリエーションを推定する手法を検討する。その発音バリエーションを発音辞書に適応した結果、発音の揺らぎやブレに対応することができ、認識精度の向上を確認できた。

キーワード 非母語音声認識, 発音辞書モデリング, 確率的発音モデル, 日本人英語

Acoustic Data-driven Pronunciation Lexicon for Non-native Speech Recognition

Satoshi TSUJIOKA[†], Liang LU^{††}, Sakriani SAKTI[†], Graham NEUBIG[†], Tomoki TODA[†], and
Satoshi NAKAMURA[†]

[†] Nara Institute of Science and Technology Takayama-cho 8916-5, Ikoma, Nara, 630-0192 Japan

^{††} Center for Speech Technology Research, University of Edinburgh, Edinburgh, EH8 9AB, UK

E-mail: †{tsujioka.satosi.tl4,ssakti,neubig,tomoki,s-nakamura}@is.naist.jp, ††liang.lu@ed.ac.jp

Abstract Nowadays, the English language is often used as a tool to facilitate communication at international meetings. Consequently, to apply automatic speech recognition (ASR) techniques for meeting dictation, the ASR system must have the capability to recognize utterances not only by native speakers but also by non-native speakers. However, the accuracy of non-native English ASR is still not satisfactory, and still has much room for improvement. To achieve this improvement, it is necessary to consider the acoustic model, pronunciation lexicon, language model, and decoding. In this study, we focus on the pronunciation lexicon. Specifically, we propose a method that first uses a G2P (Grapheme-to-phoneme) tool to predict multiple candidate pronunciations for each word, then estimates the occurrence frequency of pronunciation variations from the acoustic data of non-native speakers. In experiments, we find that the proposed method could cope with fluctuation and ambiguity of pronunciation, and it was able to achieve an improvement in recognition accuracy.

Key words Non-native speech recognition, Lexical modeling, Probabilistic pronunciation model, Japanese English

1. はじめに

音声認識技術は、計算機性能の向上及びディープニューラルネットワークを用いた音声処理技術の発展 [1] により、ノイズが少なく、話者がはっきりと標準的な発音で発話している環境下における際の精度が大きく向上している。これらの音声認識技術の進展に伴って、様々な背景雑音環境下での認識 [2]、あるいは独話・対話音声といった自然発話に対する認識 [3] も徐々に可能となりつつある。

一方、急速な国際化に伴い、英語を母語としない人々（非母語話者）が英語を話す場面が増えてきている。例えば、国際会議では、発表者は英語で発表や質疑応答を行うため、話者のほとんどが非母語話者であることは決して珍しくない。また、英語の学習支援ツールである CALL (Computer Assisted Language Learning) システム [4] での発音評定の場面を考えた場合、学習者は非母語話者であるため、正確に発音の評定をする必要がある。このような場面において、非母語話者を考慮した英語音声認識技術は必要不可欠である。

しかし、非母語話者の英語発音は英語母語話者に比べて発音の揺らぎ、ブレ、訛りといったものが原因となり、認識精度が英語母語話者に比べ低下してしまうことが多い。これらの問題に対して、音響モデル [5]、発音辞書、言語モデルなど、音声認識器の各モジュールを非母語音声に適応する必要がある。本研究では、発音辞書に焦点を当て、非母語音声の認識精度向上を図る。

英語音声認識で用いられる発音辞書を構築する際、主に人手による発音付与された辞書をもとに、表記から発音候補を予測する G2P (Grapheme-to-phoneme) ツール [7] で拡張したものが用いられる。また、G2P の誤りによる認識性能の低下を緩和するために、実発音における発音のバリエーションを予測する手法も提案されている [6]。この手法では、確率的発音モデルを用いて各単語の発音バリエーションを G2P によって生成し、実音声から生起頻度が高い発音のバリエーションを推定し、これらの発音バリエーションを音声認識用の発音辞書へ適応する。これは母語話者のための辞書作成に有用であることは先行研究により確認されている。

本研究では、この手法を参考に、非母語話者のための発音辞書作成の可能性を探る。本研究では主に二つの貢献をする。

(1) まずは、男女 100 人ずつの日本人英語音声データベース [8] の一部の音声認識を行い、発話者の英語習熟度別に実音声からの発音推定手法の有用性を検証する。

(2) 次に、あらかじめ間違えそうな発音が分かる場合、この知識を G2P ツールの学習データとなる発音辞書の作成に反映する方法を提案する。

実験の結果、非母語話者に見られる発音に適応した発音辞書生成法が従来の発音辞書を使用した時と比べ、特に、英語初級者と中級者における認識精度がそれぞれ 39.0% と 27.9% と向上した。また、本研究で提案した非母語話者用の G2P 学習データ作成法を用いた場合、それぞれの認識精度が 33.8% と 25.3% と大きく向上した。

2. 音声認識と確率的発音モデル

2.1 音声認識の定式化

従来の音声認識システムは、観測された音声特徴量を \mathbf{X} 、認識結果の単語列を $\hat{\mathbf{W}}$ とした時、以下の式で表される。

$$\hat{\mathbf{W}} = \arg \max_{\mathbf{W}} P(\mathbf{X}|\mathbf{W})P(\mathbf{W}) \quad (1)$$

式 (1) の $P(\mathbf{X}|\mathbf{W})$ は音響モデル確率、 $P(\mathbf{W})$ は言語モデル確率を表している。

さらに、各単語の音響的特徴を直接モデル化するのではなく、各単語の発音をモデル化する発音辞書を用意し、この発音に対して音響モデルを定義する。このため、式 (1) は以下の式 (2) のように書き換える。

$$\hat{\mathbf{W}} = \arg \max_{\mathbf{W}} P(\mathbf{W}) \sum_{\mathbf{B} \in \Psi_{\mathbf{W}}} P(\mathbf{X}|\mathbf{B})P(\mathbf{B}|\mathbf{W}) \quad (2)$$

ここで $\mathbf{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ は単語列 $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ の発音系列候補を表しており、各発音系列候補の確率を $P(\mathbf{B}|\mathbf{W})$ で表している。 \mathbf{b}_i は単語 \mathbf{w}_i の発音である事を示す。 $\Psi_{\mathbf{W}}$ は単語列 \mathbf{W} の考えられる全ての発音系列候補の集合を表す。^(注1) 単語の発音は前後の単語によって影響されることが多いが、ほとんどの場合はこれらの影響を無視し、各単語の発音確率を以下の式のように表す。

$$P(\mathbf{B}|\mathbf{W}) = P(\mathbf{b}_1|\mathbf{w}_1) \cdots P(\mathbf{b}_n|\mathbf{w}_n) \quad (3)$$

各単語に複数の発音系列候補がある場合、それぞれに対して発音確率を付与する。

$$P(\mathbf{b}_i = \mathbf{p}_j|\mathbf{w}_i) = \theta_{ij}, \quad j = 1, \dots, J_i \quad (4)$$

$$\sum_{j=1}^{J_i} \theta_{ij} = 1 \quad (5)$$

J_i は単語 \mathbf{w}_i の考えられる発音系列候補の数であり、 \mathbf{p}_j とは発音確率 θ_{ij} を持つ発音系列候補を指す。

2.2 確率的発音モデルの更新

発音系列候補は通常 G2P で推定されるが、その G2P の発音推定誤りは多々あり、誤った発音系列候補は音声認識に悪影響を及ぼす。この問題を解決するために、実音声から正しい発音系列候補を推定する手法が提案されている [6]。本節では Lu ら [6] の確率的発音モデルの更新について述べる。発音辞書の反復学習の仕組みを図 1 に示し、この仕組みの全体の流れについて説明する。

1. まず、小規模の発音辞書を G2P ツールの学習データとして使用し、G2P モデルの学習を行う。
2. 学習された G2P モデルから、各単語ごとに複数の発音系列候補を生成するとともに、全ての発音確率に同等の確率を付

(注1) : McGraw ら [11] の発音混合モデルを用いて、一つの単語に全ての考えられる発音系列候補を発音辞書として利用することが可能である。

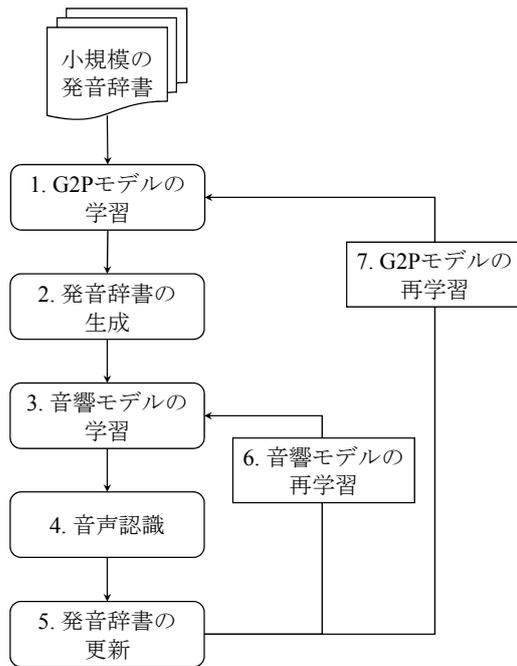


図 1 発音辞書の反復学習

Fig. 1 Lexicon iterative learning.

与する。この過程で生成された発音系列候補を *Initial* と定義し、その具体的な例を表 1 の *Initial* に示す。

3. 同等の発音確率が付与された発音辞書を用いて、音響モデルの学習を行う。
4. 対象となる実音声に対して認識を行い、認識結果の音素ラティスを取得する。
5. 認識された各単語における発音系列の音素ラティス出現回数を計算し、その単語の出現回数で割ることにより、発音確率を更新する。この際、更新された発音確率が閾値を下回った発音系列を削除する。この過程で更新された発音確率を持つ発音系列候補を *Updated* と定義し、その具体的な例を表 1 の *Updated* に示す。
6. 更新された発音確率が付与された発音辞書を使用して、再度認識を行うとともに、音響モデルの再学習を行う。
7. 5 の課程を経て再度更新された発音辞書を、G2P ツールの学習データとして使用し、G2P モデルの再学習を行う。

これらの発音辞書の反復学習から、発音確率が更新され、実音声に対して尤もらしい発音系列候補を選択することが可能となり、実音声に合わせた発音辞書を作成することができる。

3. 非母語話者のための発音辞書生成法

従来の発音辞書では、主に母語話者を想定した発音辞書を用いているため、非母語話者の発音では母語話者の発音辞書に適合していない場合が多いと考えられる。その結果、母語話者に比べると非母語話者音声の認識精度が低下してしまうことが多い。この従来の発音辞書を拡張せずに用いる場合を、これ以降、*Baseline* と呼ぶ。

この辞書の不適合の問題に対して、本研究では、前節で述べた確率的発音モデルを用いた発音辞書の反復学習を軸に四つの

表 1 発音辞書の更新の例

Table 1 Example of pronunciation learning.

Word	<i>Initial</i>		<i>Updated</i>	
	発音系列	θ	発音系列	θ
bathroom	b aa th r uw m	0.2	b ae th r uw m	1.0
	b ae th r uw m	0.2		
	b et dh r uh m	0.2		
	b ey dh r uw m	0.2		
	b ey th r uw m	0.2		
academic	ae k ah d ah m ih k	0.2	ae k ah d eh m iy k	0.58
	ae k ah d eh m ih k	0.2	ah k ah d eh m ih k	0.42
	ae k ah d eh m iy k	0.2		
	ah k ae d ah m iy k	0.2		
	ah k ah d eh m ih k	0.2		
trouble	t r ah b ah l	0.2	t r ah b ah l	0.63
	t r ah b ah l iy	0.2	t r aw b ah l	0.37
	t r ah b ah l n	0.2		
	t r aw b ah l	0.2		
	t r aw b ah l n	0.2		

発音辞書生成手法を提案し、発話者の英語習熟別に評価実験を行う。特に発音辞書推定のシードとなる G2P 学習データに着目する。

3.1 母語話者発音 G2P を用いた発音辞書生成法

まず母語話者と同様の手続きを非母語話者音声に適応する方法を考える。この場合、G2P モデルの学習データとして母語発音を想定して作られた発音辞書を用いる。具体的には、語彙数が約 4 万 3 千単語で、39 個の音素セットを用いている CMU (Carnegie Mellon University) 発音辞書を使用して生成された発音辞書である。この生成法では、母語話者用の発音辞書を G2P モデルの初期データとして用いるため母語話者に見られる発音系列候補が生成され、非母語話者で確認される形となる。このため従来の発音辞書を拡張せずに用いる場合と比較して非母語音声認識に適した発音辞書が学習されると思われる。以降では、この手法を *NativeG2P* と呼ぶ。

3.2 非母語話者発音 G2P を用いた発音辞書生成法

前節で述べた発音辞書生成法では、G2P モデルの初期学習データとして、母語話者の発音辞書を用いたが、これでは非母語話者に見られるような発音系列候補が生成されない可能性があり、実際の非母語音声に適した発音が学習されないことが考えられる。しかしながら、非母語話者の発音を反映した発音辞書を G2P モデルの初期データとして用いて、非母語話者に見られる発音系列候補を生成することができれば、非母語音声に対して、より効果的に発音が学習されると思われる。

本研究では、このような非母語話者の発音を格納した情報元として、NAD のカタカナ英語辞書 30.5^(注2) と呼ばれるコンピュータの文字入力変換辞書を用いた。この辞書には語彙総数が約 1 万 6 千単語あり、各エントリは英単語とカタカナで表現

(注2) : http://nadroom.dousetsu.com/download/download_katakana_share.html

ability アビリティ AH B IY R IY T IY
academic アカデミック AH K AH D IY M IY K UW
academic アカデミック AH K AH D EH M IY K UW

図 2 非母語発音辞書変換の例

Fig. 2 Example of converted non-native lexicon.

された日本語発音が付与されている。

これを非母語英語認識の発音辞書に変換するためにカタカナから非母語話者に見られる発音系列候補への変換を行う。カタカナから発音系列への変換には、各カタカナ文字から 3.1 節と同じ 39 個の音素セットへと変換するルールを人手で作成し、用いる。この手続きの結果から得られるエントリの具体例を図 2 に示す。

この生成法で、非母語話者に見られる発音系列候補を発音辞書に適合するため、3.1 節の母語話者発音を G2P モデルとして用いた発音辞書と比較して非母語話者の発音に近い候補が得られると思われる。以降では、この手法を Non-nativeG2P と呼ぶ。

3.3 母語・非母語話者発音辞書を結合した発音辞書生成法

3.1 節と 3.2 節で述べた発音辞書生成法では、G2P モデルの初期学習データとして母語話者と非母語話者の発音辞書を個別に用いたが、発話者の英語習熟度によって適合する発音辞書が変化する可能性がある。例えば、習熟度の高い人であれば、母語話者に近い発音話者が多くいると考えられるため、非母語話者の発音辞書のみを適合しても効果が得られない場合がある。このような習熟度依存性を回避するために、母語話者と非母語話者の両者に見られる発音候補が生成された発音辞書を組み合わせる手法を試みる。この生成法では 3.1 節と 3.2 節で生成された発音辞書を組み合わせて母語・非母語話者に見られる発音候補すべてを発音辞書に適合する。以降では、この手法を BothDict と呼ぶ。

3.4 母語・非母語話者発音 G2P を用いた発音辞書生成法

3.3 節で述べた発音辞書生成法では、二つの発音辞書を組み合わせた発音辞書生成法であり、発音辞書のエントリ数が通常の 2 倍程度となるため、各単語における発音候補の曖昧性が増加し認識精度が低下する可能性がある。最後の手法では、G2P ツールの初期学習データとして、母語話者発音辞書と非母語話者発音辞書の二つの発音辞書を用いて学習がする。これにより、発音辞書のエントリ数を前節で述べた発音辞書生成法に比べ、少量かつ両者に見られる発音系列候補が反映されると思われる。以降では、この手法を Native+Non-nativeG2P と呼ぶ。

4. 実験的評価

4.1 実験条件

本研究では Minematsu [8] による ERJ (English Read by Japanese) データベースの一部を学習・評価データに用いる。このデータベースでは、日本人が読み上げた音声に対して英語母語話者の英語教師 5 名が、(1) 音素生成、(2) リズム生成、(3) イントネーション生成の三つの観点から、1.0~5.0 の範囲でスコアリングされている。我々はこの三つのスコアリングの加算

表 2 実験データ

Table 2 Experimental data.

学習データ	人数	時間	単語数 (千)
WSJ	282	82.9	370
LOW	6	1.0	5.8
ERJ	20	3.3	16.6
HIGH	8	1.3	4.9
評価データ	人数	時間	単語数 (千)
LOW	5	0.8	4.3
ERJ	40	6.6	33.8
HIGH	20	3.3	16.6

平均を行い、発話者を三つの英語習熟度別に分割し、1.0~2.5 を LOW (初級者)、2.5~3.5 を MID (中級者)、3.5~5.0 を HIGH (上級者) とした。

認識器には Kaldi tool kit [12] を使用し、音響モデルの特徴量には 39 次元の MFCC+ Δ + $\Delta\Delta$ を用いた。学習データには音響モデル・言語モデルともに、WSJ (Wall Street Journal) と ERJ の一部を使用した。評価データには学習データに含まれていない ERJ の一部を使用した。データの詳細を表 2 に示す。また、評価基準には、WER (Word Error Rate) を使用し、この数値が低ければ認識精度が良い。比較実験としてこのデータベースの母語話者音声データ男女 10 人の認識評価実験を行った結果、WER は 8.1% となった。

母語話者の発音辞書には CMU の発音辞書約 4 万 3 千単語を使用し、非母語話者の発音辞書は NAD のカタカナ英語辞書に基づいて、カタカナを英語発音の音素に変換したものを用いた。G2P ツールとして、Bisani ら [7] の Sequitur を使用した。

4.2 実験結果

本実験結果から主に二つの項目について考察する。

(1) まずは、前節で述べた実音声からの発音推定手法が三つの英語習熟度別にどう影響されているか。

(2) 次に、どの発音辞書生成法が非母語音声の認識に有効であるかを検証する。

LOW, MID, HIGH それぞれの実験結果を図 3 に示す。横軸は反復学習のそれぞれの状態を表し、縦軸は WER を表している。従来の発音辞書を拡張せずに用いる場合の WER は、LOW では 44.5%, MID では 29.9%, HIGH では 17.5% となった。これらを Baseline として四つの発音辞書生成法の評価と分析を行う。

図 3 の結果から分かるように、LOW と MID においては、実音声からの発音推定手法を軸とした発音辞書生成法が Baseline に比べ、認識精度が向上している。NativeG2P を用いた手法では、LOW, MID それぞれの WER は 39.0%, 27.9% となった。特に、Non-nativeG2P を用いた手法では、LOW では 33.8%, MID では 25.3% と Baseline や NaiveG2P よりも大きく精度が向上し、非母語音声に見られる発音系列候補を学習した発音辞書が最も有効であることが示された。その一方、Native+Non-nativeG2P や BothDict は Non-nativeG2P を超えることはほとんどなく、母語話者発音辞書を考慮する効果が見られなかつ

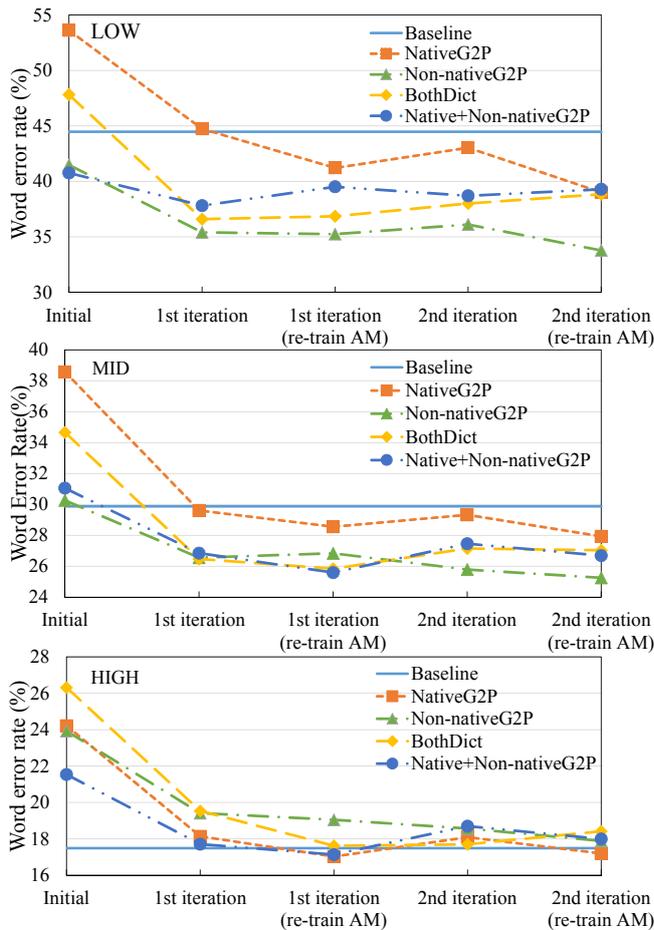


図3 LOW, MID, HIGH における評価実験結果

Fig. 3 Experiment results of LOW, MID, HIGH speakers.

た。

HIGH においては、NativeG2P と Native+Non-nativeG2P を用いた発音辞書生成法が Baseline に比べ、若干の精度向上が見られた。HIGH における発音は母語話者に近い発音話者が多くいると考えられるため、Non-nativeG2P では Baseline よりも認識精度の改善が難しかったと考えられる。

最後に Non-nativeG2P の手法を使った発音辞書の更新によって見られた従来の母語話者発音と非母語話者の発音の違いについて表 3 に示す。この例から非母語話者の実音声に見られる発音は従来の母語話者発音辞書では適合しない発音が存在することが分かる。

この三つの実験結果から G2P を用いた発音辞書学習が習熟度の低い非母語音声認識の精度向上に有効であることが示された。特に、Non-nativeG2P は、G2P モデルの初期学習に習熟度の低い非母語話者の発音を学習し、その音声に見られる発音候補を発音辞書に適応した結果、従来の発音辞書を用いた時に比べて、音声認識に有効であることが確認できた。

5. 関連研究

非母語音声の認識において、音響モデルを適応する先行研究が多い。大崎ら [13] の手法では、音響モデルの観点から日本人英語モデルと日本人日本語モデルをマルチパス化する手法と日本人英語に見られるスペルに依存した発音の癖をモデル構築に組み込

表 3 母語話者と非母語話者の発音更新の例

Table 3 Example of native and non-native pronunciation learning.

Word	母語話者発音		非母語話者発音	
	発音系列	θ	発音系列	θ
arrival	er ay v ah l	1.0	ah r ay b ow	0.67
			er ay v ah l	0.33
ballet	b ah r ey	1.0	b ah r eh ey	0.75
			b ah r eh	0.25
tibet	t ah b eh t	1.0	t ih b eh t	0.67
			t ah b eh t	0.33

む手法が提案されている。また、Imsenge ら [14] は、KL-HMM (Kullback-Leibler divergence based hidden Markov models) を用いた非母語音声のマルチアクセントに対応した音響モデル構築手法を提案している。本研究では発音辞書に焦点を当てたが、上記の音響モデル更新法と合わせて利用できると考えられる。

非母語音声認識の発音辞書における関連研究において、Pongkittiphan ら [16] の手法では、DTW (Dynamic Time Warping) を用いた日本人英語とアメリカ人英語との IPA 書き起こし距離に基づく発音距離予測を使用することで英語母語話者が聞き取るのが難しいような日本人英語の不明瞭なアクセントに対して、高精度の発音予測をすることで、非母語音声認識精度の向上に活用できる手法を提案している。この手法では、主に日本人英語とアメリカ人英語の発音距離に着目した手法である。Razavi ら [15] の手法では、本研究でも用いている確率的発音モデルと、[14] の KL-HMM と hybrid HMM/ANN (artificial neural network) を用いた手法で、非英語母語音声に対して有用な発音系列候補を生成することで認識精度を向上を確認している。Lehr ら [17] の手法では、識別的言語モデルを用いた識別的発音モデル学習を提案している。この手法ではテキスト情報を用いた発音変換ルールを知識データベースとして作成し、英語表記から非母語音声に見られる発音に変換して識別的発音モデルを作成している。Rasipuram ら [18] の手法では、本研究と同じように確率的発音モデルを音声認識の発音辞書学習に使用し、この手法では本研究では用いていない単語表記列だけの情報から発音予測フレームワークを用いた手法を提案している。これらの手法では、実際に非母語話者の発音辞書を生成しておらず、正確な非母語話者に見られる発音候補を発音辞書に適応できない手法である。本研究では、非母語話者発音辞書を生成するとともに、実音声からの発音推定手法を発音辞書に適応しているため、実際の非母語話者に見られる発音候補を選択し、発音辞書に反映することが可能である。

6. ま と め

本稿では、非母語音声の認識精度改善のために、確率的発音モデルとその反復学習による実音声を用いた発音辞書獲得の手法を提案した。また、非母語話者発音を G2P モデルとして用いた発音辞書を使用し、あらかじめ間違えそうな発音を非母語話者に適応した発音辞書生成法を提案した。実験的評価結果から、提案手法の非母語話者発音を G2P モデルとして用いた発音辞書生成法が、従来の母語話者の発音辞書を用いた時に比べ、非母語音声認識の精度向上が確認でき、提案手法の有用性が示された。

今後は、本稿の提案手法を複数の非英語母語音声の認識に適用する手法や、英語母語話者の発音から非母語話者に見られる発音の癖を考慮して母語話者の発音から部分的に変換し、発音辞書に適用する手法を検討する。

謝辞

本研究の一部は、JSPS 科研費 24240032 および 26870371 の助成を受け実施した。

文 献

- [1] 河原達也, "音声認識の方法論に関する考察—世代交代に向けて—", 情報処理学会研究報告, SLP2014-100, Jan. 2014.
- [2] J.Li, L.Deng, Y.Gong, R.Haeb-Umbach, "An Overview of Noise-Robust Automatic Speech Recognition," IEEE/ACM TASLP, VOL. 22, NO. 4, pp.745–777, April. 2014.
- [3] C.Wang, S.Senef, "Lexical Stress Modeling for Improved Speech Recognition of Spontaneous Telephone Speech in the JUPITER Domain," in Proc. INTERSPEECH 2001, 2001.
- [4] 河原達也, 峯松信明, "音声情報処理技術を用いた外国語学習支援", 電子情報通信学会論文誌, Vol. J96-D No. 7, pp. 1549–1565, 2013.
- [5] T.Fraga-Silva, J.Luc Gauvain, L.Lamel, "Speech Recognition of Multiple Accented English Data Using Acoustic Model Interpolation," Signal Processing Conference (EU-SIPCO), 2014 Proceedings of the 22nd European. IEEE, pp. 1781–1785, Sept. 2014.
- [6] L.Lu, A.Ghosal, S.Renals, "Acoustic Data-driven Pronunciation Lexicon for Large Vocabulary Speech Recognition," in Proc. ASRU. IEEE, 2013.
- [7] M. Bisani, H. Ney, "Joint-sequence Models for Grapheme-to-phoneme Conversion," Speech Communication, vol. 50, no.5, pp. 434–451, 2008.
- [8] N.Minematsu, et al., "English Speech Database Read by Japanese Learners for CALL System Development," in Proc.LREC2002, pp.896–903, 2002
- [9] F.Beaufays, A.Sankar, S.Williams, M.Weintraub, "Learning Name Pronunciations in Automatic Speech Recognition Systems," in Proc. ICTAL. IEEE, 2003, pp. 233–240.
- [10] X.Li, A.Gunawardana, A.Acero, "Adapting Grapheme-to-phoneme Conversion for Name Recognition," in Proc. ASRU. IEEE, 2007, pp. 130–135.
- [11] I.McGraw, I.Badr, J.Glass, "Learning Lexicons From Speech Using a Pronunciation Mixture Model," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 2, pp.357–366, 2013.
- [12] D.Povey, A.Ghoshal, G.Boulianne, L.Burget, O.Glembek, N.Goel, M.Hannemann, P.Motlicek, Y.Qian, P.Schwarz, J.Silovsk´y, G.Semmer, K.Vesel´y, "The Kaldi Speech Recognition Toolkit," in Proc. ASRU, 2011.
- [13] 大崎功一, 峯松信明, 広瀬啓吉, "日本人英語発声に観測される発音上の癖を考慮した音声認識", 電子情報通信学会技術研究報告, SP2002-180, Mar. 2003.
- [14] D.Im seng, R.Rasipuram, M.Magimai.-Doss, "Fast and Flexible Kullback-Leibler Divergence Based Acoustic Modeling for Non-native Speech Recognition," in Proceedings of ASRU, pp. 348–353, Dec. 2011.
- [15] M.Razavi, M.Magimai Doss, "On Recognition of Non-Native Speech Using Probabilistic Lexical Model," in Proc. INTERSPEECH 2014, Sept. 2014.
- [16] T.Pongkittiphon, N.Minematsu, T.Makino, K.Hirose, "Improvement of Intelligibility Prediction of Spoken Word in Japanese Accented English Using Phonetic Pronunciation Distance and Word Confusability," in Proc. O-COCOSDA, pp. 276–281, Sept. 2014
- [17] M.Lehr, K.Gorman, I.Shafran, "Discriminative Pronunciation Modeling for Dialectal Speech Recognition," in Proc. INTERSPEECH 2014, pp. 1458–1562, Sept. 2014.
- [18] R.Rasipuram, M.Razavi, M.Magimai-Doss, "Integrated Pronunciation Learning for Automatic Speech Recognition Using Probabilistic Lexical Modeling," in Proc. ICASSP 2015, 2015.