Vol.2015-SLP-109 No.14
2015/12/2

発音変換知識を用いないデータ駆動型発音学習による非母語話者の音声認識

辻岡　聡[1,a)]　サクティ　サクリアニ[1]　ニュービッグ　グラム[1]　吉野　幸一郎[1]　中村　哲[1]

概要：急速な国際化に伴い，国際会議などでは英語が国際共通語として用いられ，非英語母語話者の間でも英語で意思疎通を図る場面が増加している．このような非母語音声を認識して会議録を作成するなどの応用技術を考えた場合，非母語音声認識を高精度に行う必要がある．しかし，非英語母語話者の音声は英語母語話者に比べ，発音の揺らぎ，ブレ，訛りが存在する．そのため，非母語音声の音声認識精度は母語音声よりも低下することが多い．非母語音声認識における非母語音声の影響は，音響モデル，発音辞書，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．非母語音声認識において，人手による発音変換知識を用いた非英語母語話者発音の逐次学習が提案されている．しかし，これらの発音変換知識を人手で構築することは困難であり，汎用性が小さい．そこで本研究では，非母語音声の音素認識結果を G2P (Grapheme-to-phoneme) ツールの学習データとして用いる．さらに，G2P ツールから非英語母語話者に見られる複数の発音バリエーションを生成し，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提案する．その結果，人手による発音変換知識を用いた時と比べ，英語中級者においてはほぼ同等の精度，英語上級者においては約 2.4 %の精度向上を確認できた．

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

Non-native ASR Utilizing Acoustic Data-driven Pronunciation Learning with Zero Knowledge of Non-native Pronunciation

Satoshi Tsujioka[1,a)]　Sakriani Sakti[1]　Graham Neubig[1]　Koichiro Yoshino[1]
Satoshi Nakamura[1]

Abstract: Non-native speech differs significantly from native speech, often resulting in a degradation of the performance of automatic speech recognition (ASR). Handcrafted pronunciation lexicons used in standard ASR systems generally fail to cover non-native pronunciations, and design of new ones by linguistic experts is time consuming and costly. A previous study proposed a method to automatically learn a pronunciation lexicon in an iterative fashion using knowledge of non-native pronunciation. However, this previous method needs a handcrafted non-native pronunciation lexicon to train a grapheme-to-phoneme (G2P) converter used to generate non-native pronunciation variations, including pronunciations of new words. This non-native pronunciation lexicon is difficult to obtain, and lacks versatility to be applied to other non-native speakers. This study proposes a method for non-native ASR using acoustic evidence for pronunciation learning without knowledge of non-native pronunciation. In experiments, we evaluate our ASR systems for speakers with three degrees of English proficiency level. The results reveal that the proposed method can achieve almost same recognition accuracy with a system using knowledge of a non-native pronunciation, and is able to achieve an improvement of about 2.4% in recognition accuracy, particularly for high-proficiency speakers.

Keywords: Non-native speech recognition, Lexical modeling, Probabilistic pronunciation model, Japanese English

1. はじめに

計算機の性能向上や深層学習の登場により音声処理技術が発展 [1] している．雑音が少なく，発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いては音声認識精度が大きく向上しており，会議の議事録などの実用的な場面で音声認識が使われ始めている．一方，急速な国際化に伴い，英語を母語としない人々（非英語母語話者）が共通語として英語を話す場面が増えてきている．例えば国際会議では，様々な国から集まった発表者が英語で発表や質疑応答を行うため，話者の多くが非英語母語話者であることは珍しくない．また，国際化に伴い非英語母語話者が英語能力を身につけることが必要不可欠な状況になってきており，英語の学習支援ツールである CALL (Computer Assisted Language Learning) システム [2] が開発されている．CALL システムでは音声認識技術を用いて，学習者である非英語母語話者の発音評定を行う．そのため，正確に非英語母語話者の発音を認識する必要がある．このような場面において，非英語母語話者を考慮した英語音声認識技術の精度向上は必要である．

非英語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存在する．その結果，非英語母語話者の英語音声認識精度が英語母語話者に比べて低下してしまうことが多い．そのため，音響モデル [3]，発音辞書 [4] などの音声認識器の各モジュールを非英語母語話者の発話に適応させる必要がある．そこで本研究では，発音辞書に焦点を当て非英語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 [5] を拡張したものが用いられる．この G2P では，Lu ら [6] によって G2P の誤りによる認識性能の低下を緩和するために以下の三つの枠組みによる手法が提案されている．まず，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 によって生成する．次に，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．最後に，これらの発音バリエーションによって音声認識用の発音辞書を適応させる．これは英語母語話者のための発音辞書生成に有用であることが先行研究により確認されている．

我々は，この手法を参考にした非英語母語話者のための発音辞書生成法を提案 [7] しており，G2P の学習データとして人手による発音変換知識を用いた非英語母語話者発音を用いることで，非英語母語話者に見られる潜在的な発音候補を生成し，[6] の手法を適用した発音辞書生成を行った．その結果，英語初級者・中級者において非英語母語話者発音を G2P の学習データとして用いた発音辞書生成法が有効であることが確認されている．

[7] の手法では，G2P の学習データとして，あらかじめ非英語母語話者の発音変換知識が必要である．しかし，Besacier ら [8] より，非英語母語音声や発音変換知識を収集するのは困難であり，それらのデータは低資源であることが確認されている．そのため，先行研究で用いるような発音変換知識を人手で構築することは困難であり，他の非母語音声認識への応用が難しく，汎用性が小さい．そこで本研究では，この手法 [7] を発展させ，発音変換知識を用いない非英語母語話者のための発音辞書生成法の可能性を探る．本研究の主な貢献は次の二つである．まず，日本人学生英語音声データベース [9] の一部の音声認識を行い，発音変換知識を用いない実音声からの発音推定手法の有用性を検証する．次に，[7] の人手による発音変換知識を用いた発音辞書生成法との認識精度の比較を行い，本稿の提案手法の有用性を検証する．また，音響モデルの代表的な適応手法である話者適応学習 (Speaker Adaptive Training: SAT) との比較も行う．

実験の結果，人手による発音変換知識を用いた時と比べ，英語中級者においてはほぼ同等の認識精度，英語上級者においては約 2.4 %の認識精度向上を確認できた．

2. 関連研究

非母語音声の認識において，音響モデルを適応した先行研究がいくつかある．Wang ら [10] は，PDTS (Polyphone Decision Tree Specialization) と呼ばれる決定木学習を用い，二つ以上の発音を持つ文字に対するクラスタリングを行うことで，前後の単語によって変化する英語母語話者発音と非英語母語話者発音とのミスマッチに対応する手法を提案している．大崎ら [11] は，日本人の英語発音に見られる英語音素と日本語音素との置換分析を行い，この分析結果を用いて発音が類似する英語音素と日本語音素との対応関係を音響モデルのマルチパス化へ適用する手法と，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む手法を提案している．Imseng ら [12] は，KL-HMM (Kullback-Leibler divergence based Hidden Markov Models) を用いることで，従来の GMM(Gaussian Mixture Model)-HMM で使用されるパラメータ数よりも少ないパラメータ数かつ，少量の非母語音声データから学習できる手法を提案している．本研究で提案する発音辞書の改善は，これらの音響モデル適応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 [13] は，DTW(Dynamic Time Warping) を用いた日本人英語とアメリカ人英語それぞれの IPA 書

[1] 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), Japan
[a)] tsujioka.satoshi.tl4@is.naist.jp

き起こし距離に基づく発音距離予測を使用している．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発音に対して，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．Lehr ら [14] 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生成する手法を提案している．Rasipuram ら [15] は，確率的発音モデルと [12] の両者の手法を用いて発音学習に使用している．この際 G2P を用いるのではなく，文字表記に基づく音素を扱う音声認識フレームワーク内で発音予測をする手法を提案している．

これに対し本研究では，非英語母語話者の発音変換知識を音素認識によって補い，G2P を用いて非英語母語話者の潜在的な発音候補を生成する．また，確率的発音モデルを用いることで効率的に発音辞書を実音声に対して適応した．これにより，実際の非英語母語話者に見られる発音候補を発音辞書に反映することが可能である．

3. 音声認識と確率的発音モデル

3.1 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．

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} P(\mathbf{X}|\mathbf{W})P(\mathbf{W}) \tag{1}$$

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．ここで，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．

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} P(\mathbf{W}) \sum_{B \in \Psi_{\mathbf{W}}} P(\mathbf{X}|\mathbf{B})P(\mathbf{B}|\mathbf{W}) \tag{2}$$

ここで $\mathbf{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ は単語列 $\mathbf{W} = \{\mathbf{w}_1, ..., \mathbf{w}_n\}$ の発音系列候補を表しており，単語列に対する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．$\Psi_{\mathbf{W}}$ は単語列 $\mathbf{W}$ の考えうる全ての発音系列候補の集合を表す．[*1] 今回は各単語の発音は当該単語のみに依存すると仮定し，各単語の発音確率を以下の式のように表す．

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

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．

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

[*1] McGraw ら [16] 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．

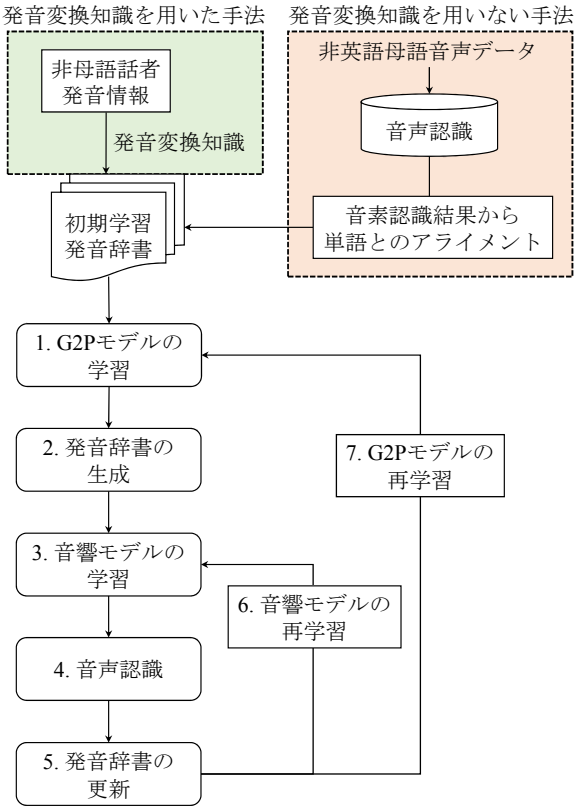

図 1 実音声を用いた逐次発音学習
Fig. 1 Acoustic data-driven lexicon iterative learning.

$$\sum_{j=1}^{J_i} \theta_{ij} = 1 \tag{5}$$

J_i は単語 $\mathbf{w}_i$ の考えう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．

3.2 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，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]．この発音辞書の逐次学習の仕組みを図 1 に示し，以下に全体の流れを説明する．

1. 初期学習発音辞書を 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. 等確率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. 学習データの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逐次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．

4. 非英語母語話者のための発音辞書生成法

従来の発音辞書では，主に英語母語話者を想定した発音辞書を用いている．そのため，非英語母語話者の発音と発音辞書のミスマッチが問題となり，英語母語話者に比べると非英語母語話者の音声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，Baseline とする．

この発音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逐次学習にもとづく発音辞書生成手法を提案する．この際，特に発音辞書推定のシードとなる G2P 学習データに着目し，図 1 に見られる二つの発音辞書生成法について述べる．

4.1 非英語母語発音変換知識を用いた発音辞書生成法

本手法は先行研究 [7] の発音辞書生成法の一つである．非英語母語話者の発音変換知識を反映した発音辞書を生成するため，G2P モデルの学習データに NAD のカタカナ英語辞書 30.5[*2] を元に変換した発音辞書を用いている．本辞書を用いた理由は，日本人に見られる英語発音がカタカナ発音に準ずる傾向があると考え，本辞書を用いることで日本人に見られる英語発音を効果的に学習できると考えたからである．NAD のカタカナ英語辞書 30.5 は，コンピュータの文字入力変換辞書であり，語彙総数は約 1 万 6 千単語で，各エントリは英単語とカタカナで表現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，カタカナから非英語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英語発音に見られる 39 個の英語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．この手法を KnowledgeG2P と呼ぶ．

[*2] http://nadroom.dousetsu.com/download/download_katakana_share.html

表 1 発音辞書の更新の例
Table 1 Example of pronunciaton learning.

Word	*Initial* 発音系列	θ	*Updated* 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		

ability | アビリティ | AH B IY R IY T IY
academic | アカディミック | 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.

4.2 非英語母語発音変換知識を用いない発音辞書生成法

前節で述べた発音辞書生成法では，G2P モデルの学習データとして，発音変換知識にもとづく非英語母語話者の発音辞書を用いた．しかし，発話者の英語習熟度によって適応した発音辞書が変化する可能性がある．例えば，発話のなかに英語母語話者に近い発音が混在しているようなユーザの場合，前節の発音辞書生成法のみでは適切な対応する発音が得られない場合がある．これに対応するためには，あらかじめその非英語母語話者の発音変換知識が必要となるが，人手でその発音変換知識を構築することはコストが高く，他の非母語音声認識へ応用できる汎用性も小さい．

そこで本手法では，非母語音声の音素認識結果から単語とのアライメントを取ったものを G2P の学習データとして用いている．そのため，人手による発音変換知識が無くても G2P を用いて非英語母語話者の潜在的な発音候補を生成することが可能である．また，他の非母語音声にもこの辞書生成法で適応させることが可能だと考えられる．この手法を No-KnowledgeG2P と呼ぶ．

5. 実験的評価

5.1 実験条件

本研究では Minematsu[9] による ERJ (English Read by Japanese) データベースの一部を学習・評価に用いる．こ

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	MID	93	3.3	72.4
	HIGH	26	1.3	18.0
評価データ		**人数**	**時間**	**単語数 (千)**
	LOW	5	0.8	4.3
ERJ	MID	40	6.6	33.8
	HIGH	20	3.3	16.6

のデータベースでは，日本人学生が読み上げた英語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している．我々はこの三つのスコアリングの加算平均を行い，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．

音声認識器は Kaldi tool kit[17] を使用し，音響モデルの特徴量は 39 次元の MFCC+Δ＋ΔΔ を用いている．また，線形判別分析 (Linear Discrinative Analysis: LDA) と最尤線形変換 (Maxium Likelihood Linear Transform: MLLT) を用いた特徴量変換にもとづく次元数圧縮を行っており，対象フレームの前後 3 フレーム (計 7 フレーム) を考慮した音響モデル学習を行っている．学習データは音響モデル・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．これらのデータの詳細を表 2 に示す．評価基準は単語誤り率 (Word Error Rate: WER) を用いる．

英語母語話者の発音辞書には CMU 発音辞書 [*3] を使用している．非英語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 ツールは，Bisani ら [5] の SequiturG2P を使用した．

5.2　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) 音素認識結果を G2P の学習データとして用いた発音辞書生成法と，非母語発音変換知識を用いた発音辞書生成法との認識性能を比較し，その有効性を検証する．
(2) 次に，音響モデルの代表的な適応手法である特徴量空間最尤線形回帰 (feature-space Maximum Likelihood Linear Regression: fMLLR) を用いた話者適応学習との認識性能を比較し，検証する．

LOW，MID，HIGH それぞれの実験結果を図 3 に示す．

[*3] http://svn.code.sf.net/p/cmusphinx/code/trunk/cmudict/

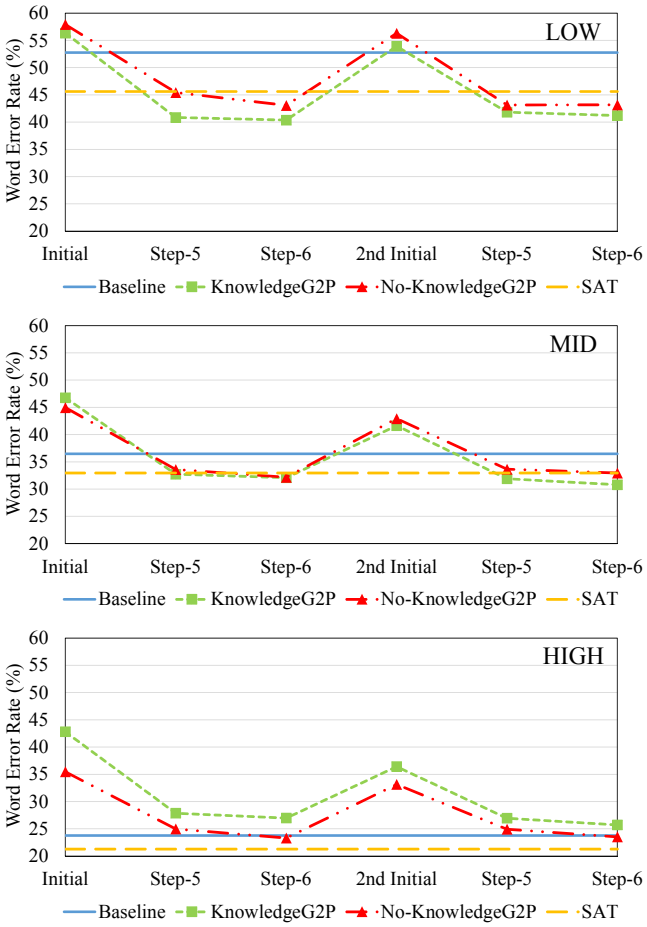

図 3　LOW, MID, HIGH における評価実験結果
Fig. 3　Experiment results of LOW, MID, HIGH speakers.

縦軸は WER を表しており，横軸は逐次学習のそれぞれの状態を表している．Initial は各単語の発音候補に対して等確率の発音確率が付与された状態での認識精度，Step-5 は更新された発音確率を用いて認識を行った際の認識精度，Step-6 は音響モデルの再学習を行った際の認識精度を表している．従来の発音辞書を拡張しない場合の誤り率は，LOW では 52.8%，MID では 36.5%，HIGH では 23.8%となった (Baseline)．これらを基準として，提案手法の評価と分析を行う．また，比較実験として，話者適応学習を用いた認識評価実験を行った結果，LOW では 45.6%，MID では 33.0%，HIGH では 21.3%となった (SAT)．

まず図 3 の結果から，LOW と MID においては，KnowledgeG2P と No-KnowledgeG2P の両者が Baseline と比較して，認識精度が向上していることが確認できた．KnowledgeG2P を用いた発音辞書生成法の誤り率が，LOW では 40.4%，MID では 30.8%と Baseline や SAT よりも精度が向上し，非英語母語話者の発音変換知識を用いた発音逐次学習の有効性を確認できた．また，No-KnowledgeG2P での認識精度が KnowledgeG2P とほぼ同等の認識精度を出力していることから，音素認識結果を用いた手法も同等に有効であることが確認できた．HIGH においては，話者適応学習の誤り率が最も低かった．また，No-KnowledgeG2P

表 3　英語母語話者と非英語母語話者の発音更新の例
Table 3 Example of native and non-native pronunciaton learning.

Word	英語母語話者発音		非英語母語話者発音	
	発音系列	θ	発音系列	θ
asia	ey zh ah	1.0	ey zh ah	0.44
			ey sh ah	0.56
tiphook	t ih p hh uw k	1.0	t ih p hh uw k	0.33
			t ih p hh ow k	0.67
tibet	t ah b eh t	1.0	t ih b eh t	0.67
			t ah b eh t	0.33
why	hh w ay	0.5	hh w ay	0.19
	w ay	0.5	w ay	0.81

を用いた発音辞書生成法の誤り率が 23.3%を示し，KnowledgeG2P と比較して，約 2.4%の認識精度向上が見られた．これは，HIGH における発音は英語母語話者に近い発音が混在している話者が多く存在し，KnowledgeG2P では認識精度の改善が難しかったと考えられる．

まとめると，音素認識結果を用いた発音逐次学習が KnowledgeG2P とほぼ同等の性能を示した．特に英語上級者においては，KnowledgeG2P と比較して認識精度の向上が確認できた．これらの結果から，No-KnowledgeG2P が非母語音声認識において有効であることが示された．

最後に No-KnowledgeG2P を用いた発音辞書の更新によって得られた非英語母語話者の発音と，対応する単語の英語母語話者発音の違いについて表 3 に示す．この例から，非英語母語話者の実音声に見られる発音は，従来の英語母語話者発音と大きく異なることが確認できる．

6. まとめ

本稿では，非母語音声の認識精度改善のために，非英語母語話者の音素認識結果を用いた発音逐次学習による発音辞書生成法を提案した．実験的評価結果から，提案手法が先行研究 [7] の発音辞書生成法と比較して，英語中級者においてはほぼ同等の精度，英語上級者においては約 2.4 %の精度向上を確認できた．提案手法は先行研究 [7] と比較して，非母語話者の発音変換知識が不要である点と，他の非英語母語話者への適用が容易であるという利点がある．このことから，提案手法が非母語音声認識において非常に効果的であるといえる．

今後は，本稿の提案手法を用いて日本人以外の複数の非英語母語音声の認識に適用する手法や，英語習熟度別に発音辞書を生成し，話者の英語習熟度に適応した発音辞書を自動で選択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．

参考文献

[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] 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.
[3] T.Fraga-Silva, J.Luc Gauvain, L.Lamel，”Speech Recognition of Multiple Accented English Data Using Acoustic Model Interpolation,”，Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European. IEEE, pp. 1781–1785, Sept. 2014.
[4] Raux, Antoine. ”Automated lexical adaptation and speaker clustering based on pronunciation habits for non-native speech recognition.” Interspeech. 2004.
[5] M. Bisani，H. Ney, ”Joint-sequence Models for Grapheme-to-phoneme Conversion,”，Speech Communication, vol. 50, no.5, pp. 434–451, 2008.
[6] L.Lu，A.Ghosal, S.Renals, ”Acoustic Data-driven Pronunciation Lexicon for Large Vocabulary Speech Recognition,”，in Proc. ASRU. IEEE, 2013.
[7] 辻岡聡, ルーリアン, and サクティサクリアニ. ”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 (音声).” 電子情報通信学会技術研究報告= IEICE technical report: 信学技報 115.146 (2015): 1-6.
[8] Besacier, Laurent, et al. ”Automatic speech recognition for under-resourced languages: A survey.” Speech Communication 56 (2014): 85-100.
[9] N.Minematsu, et al., ”English Speech Database Read by Japanese Learners for CALL System Development,”, in Proc.LREC2002, pp.896–903, 2002
[10] Wang, Zhirong, Tanja Schultz, and Alex Waibel. ”Comparison of acoustic model adaptation techniques on non-native speech.” Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP’03). 2003 IEEE International Conference on. Vol. 1. IEEE, 2003.
[11] 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.
[12] D.Imseng, 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.
[13] T.Pongkittiphan, 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
[14] M.Lehr, K.Gorman, I.Shafran, ”Discriminative Pronunciation Modeling for Dialectal Speech Recognition,”, in Proc. INTERSPEECH 2014, pp. 1458–1562, Sept. 2014.
[15] R.Rasipuram, M.Razavi, M.Magimai-Doss, ”Integrated Pronunciation Learning for Automatic Speech Recognition Using Probabilistic Lexical Modeling,”, in Proc. ICASSP 2015, 2015.
[16] 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.
[17] 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.