英語の発音指導に対する音声認識エンジンの利用可能性の調査

小泉明日香[*1]・岡井光輝[*2]・大野恵理[*3]・須曽野仁志[*3]・高瀬治彦[*4]・北英彦[*4]
Email: kita_yh@yahoo.co.jp

*1：三重大学工学部電気電子工学科
*2：奈良先端科学技術大学院大学先端科学技術研究科先端科学技術専攻情報理工学プログラム
*3：三重大学教育学部
*4：三重大学工学部研究科

◎Key Words　　フォニックス，音声認識，個別学習

1.　はじめに

2020 年より実施された新学習指導要項において，小学校における外国語（英語）学習において，英語の音声から文字への学習に円滑に接続することは困難であるという課題が挙げられている(1).

この課題を解決する英語の発音指導方法として広く用いられているものの 1 つにフォニックスがある．フォニックスとは，英語の文字と綴りの読み方のルールであり，英語圏の子どもたちに読み書きを教えるために開発されたものである(2).

三重大学の工学部津と教育学部との学内の共同研究として，iPad 用アプリケーション「Let's Phonics!!」を 2017 年度に開発を始め 2021 年度に当初予定していた機能の実装を完了した．図 1 に学習画面を示す．このアプリケーションは，段階的学習機能やテスト機能に加えて録音・再生機能が備わっており，発音学習が可能である(3). しかし，発音の評価機能を備えていないために，児童および英語が得意ではない先生が，客観的に発音の良し悪しを判断できないという問題がある．

日本人の英語学習者に対する発音指導方法には，フォニックスのほかに，プロソディーを中心とした指導法がある．プロソディーとは，自然な英語の発話の中における，言語の音の特徴を表す総称である．その代表例が，

① リズム（緩急・間・テンポ）
② ストレス（強弱・強勢・アクセント）
③ イントネーション（高低・抑揚）

である．日本人なまりの英語の発音に対しては，この指導法が有効であると言われている(4).

プロソディー重視の指導法は，対面指導が一般的であるが，大学生を被験者として，音声認識エンジンを採用した英語発音評価ソフトウェアを用いた指導を試みた実験の例がある(5). しかし，音声認識による英語の発音評価の正確性や妥当性は検証されていない．

近年，人工知能（AI）の技術が進歩したことで，精度を増した音声認識の分野が注目されている．音声認識は，人の話し言葉や音声を分析し，文字に変換したり機器を操作したりする技術のことで，様々なソフトウェアや製品の開発に応用されている．

例えば，iPhone に搭載された「Siri」(6)をはじめとする音声アシスタントや，Amazon で開発された「Alexa」(7)のような，住宅で行うあらゆる操作を音声指示で自動化する機器などがあげられる．

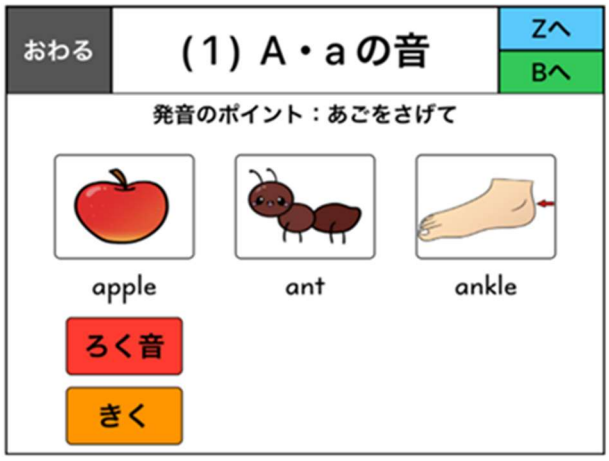

図 1 「Let's Phonics!!」発音学習画面

これらに代表される音声認識の機能は，他の開発者がソフトウェアなどを開発するうえで，共通の機能を利用するためのインタフェースとして，ほとんどの音声認識アプリケーションは Application Programming Interface（以降，API）を公開している．

2.　研究目的

「Let's Phonics!!」において，児童および英語が得意ではない先生が，客観的に発音の良し悪しを判断できないという問題を，音声認識の機能を用いて解決する手段を検討する．しかし，音声認識は話者の発音を認識することが目的であり英語の発音評価に特化して開発された機能ではない．英語の発音評価に対して音声認識を利用した際の正確性・妥当性は未検証であるため，本研究では利用可能性を調査する．

3.　研究手法

Apple が提供している API の「Speech Framework」を音声認識エンジンのサーバー版を調査対象とする．これは，「Let's Phonics!!」が iOS のタブレットアプリケーションとして開発されていることと，音声認識の使用制限が少なく，無料で提供されていることを加味している．さらに，オンデバイス版と比較したときに，音声認識の認識率が高いと予想したためである．

ネイティブの発音とノンネイティブの発音をテストデータとして収集し，音声認識させた出力結果から英語の発音評価への利用可能性を検証する．

3.1 録音アプリケーション

音声のテストデータを収集するため，専用の録音アプリケーションを作成した．音声のテストデータ収集に協力していただく被験者用に仕様を決定した．被験者としては，ネイティブスピーカーとして外国人指導助手と，「Let's Phonics!!」の使用者である小中学校の児童・生徒を想定している．特に児童のために，録音する単語の綴りが読むことができなくても利用できるように工夫した．

本アプリは被験者に合わせて，日本語と英語の二言語に対応している．被験者は，図 2，図 3 に示す画面で言語を選択し名前を入力する．これは，複数の被験者の音声データを整理するための仕様であり，本名は非公表とする．

必要となる音声のテストデータは，単語・文の単位で収集する．テストデータ項目の内容については大野に依頼した．「r」などをはじめとする，日本人が苦手とする発音を含む英単語，フォニックスの学習において使用する英単語を含む「a」から「z」までの全 26 単語，小学校 4 年生から 6 年生までのレベル別に全 33 フレーズを用意した．図 4，図 5 に音声のテストデータ項目の表示画面の例を示す．被験者は録音する単語またはフレーズを選択する．

そして，図 6 に示す録音画面で音声を録音する．録音するテストデータの項目を被験者に示すため，単語またはフレーズを画面上部に表示させる仕様にした．その下の画像は，録音するテストデータの項目に合わせたイラストを表示したボタンであり，押すと見本音声を聞くことができる．マイクのボタンを押して録音することができる．

図 2　言語選択・名前入力画面（日本語版）

図 3　言語選択・名前入力画面（英語版）

図 4　テストデータ項目（単語）表示画面

図 5　テストデータ項目（フレーズ）表示画面

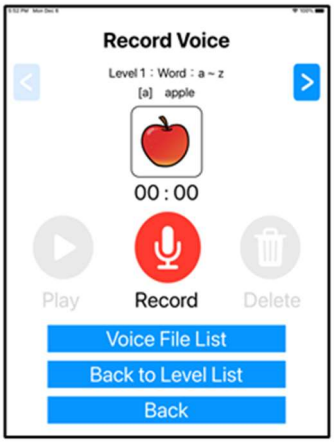

図 6　録音画面

3.2 被験者

今回の調査に協力していただいた被験者は以下のとおりである．

① 見本音声（D さん）：1 名
② ネイティブ（外国人指導助手）：4 名
- A さん（アメリカ人）
- B さん（ラテン系アメリカ人）
- C さん（中東出身カナダ人）
- D さん（カリブ諸国出身）

③ ノンネイティブ：2 名
- E さん（日本人）
- F さん（日本人）

ネイティブの発音データを収集するために，外国語指導助手 4 名に協力を依頼した．ノンネイティブの発音データの収集においては，計算機工学研究室の日本人 2 名に協力を依頼した．日本人の方には，ネイティブの発音は意識せず，日本人なまりの英語（カタカナ英語）の発音で音声を録音するように指示をした．

録音アプリケーションの見本音声は，外国人指導助手の D さんに担当していただいている．D さんの音声において，①では見本らしくゆっくりとした発音，②では見本を意識しない自然な発音を録音しているため，被験者は同じであるが，区別して音声認識させた．

3.3 音声認識

Apple が提供している音声認識 API の「Speech Framework」を用いて調査を行った．これは，「Let's Phonics!!」が iOS のタブレットアプリケーションとして開発されていることと，音声認識の使用制限が少なく，無料で提供されていることを加味している．音声認識を行う際，サーバーとの通信を行う方法とオンデバイスで行う方法の 2 通りあるが，サーバーを介した音声認識の方が，オンデバイスよりも認識率の精度が上がると予想されるため，サーバー版を利用した．

収集した音声のテストデータそれぞれに対して，北米英語で音声認識をさせて，出力結果の比較・検討を行い，英語の発音評価に対して音声認識を利用した際の正確性・妥当性を検証した．

音声認識を行った例（A さん：単語「bear」）と，出力結果として表示できる項目の例を表 1 に示す．認識結果の最有力候補として「bestTranscription」が出力される．他候補があれば「alternativeSubstrings」，それぞれの認識結果の信頼度が出力される．表 1 において，単語間の平均休止時間が「-1」秒として出力されているのは，音声データが 1 単語であったためである．

英語の発音において，ただひとつの正解というものは存在しない．音声認識を行った際，ネイティブ（英語を母国語とする人）が聞き取ることのできる発音に対して，音声認識の認識率が高くなること，そして，日本人なまり（カタカナ英語）の発音に対して，音声認識の認識率が低くなることの 2 点が望まれる．

表 1　音声認識結果の例（A さん：単語「bear」）

bestTranscription	segments	
formattedString	substring	alternativeSubstrings
文字列（全体）	文字列（一部）	他候補
Bear	Bear	
		Beer
		Better
segments		
timestamp	duration	confidence
発話開始時間［秒］	発話時間［秒］	信頼度［%］
0.48	0.6	83.4
		1.4
		1.2
speechRecognitionMetadata		
speakingRate		averagePauseDuration
1分間あたりの話す速度［単語数 / 1分］		単語間の平均休止時間［秒］
100		-1

4.　研究結果

収集した音声のテストデータそれぞれに対して，北米英語で音声認識を行った結果のまとめを表したものが，表 2，表 3，表 4 である．

表 2，表 3 において，「第一候補」の数は，表 1 のように，発音した単語・フレーズと認識結果の最有力候補が一致した数を表している．そのうちで「信頼度 50%未満」の数と，「信頼度 50%以上」の数を表している．

「信頼度 50%以上」の数を比較すると，ネイティブの発音の方がノンネイティブ（日本人）の発音よりも，音声認識を行った際の結果が良いことが分かる．そして，単語の認識結果よりも，フレーズの認識結果が良いことが，ネイティブの発音とノンネイティブの発音の認識結果で共通している．

フレーズの音声認識は，発音した前後の単語から推測することにより出力結果の信頼度を高めている．この事実は，表 3 のノンネイティブの発音の認識結果に着目すると，単語の認識結果における「信頼度 50%以上」の数が少ないことにもかかわらず，フレーズの認識結果における「信頼度 50%以上」の数が多くなっていることから分かる．特に F さんの認識結果は顕著である．

しかし，フレーズの認識結果の信頼度が高いことは，必ずしも英語の発音の正確性が高いことを表していないという点が重要である．

表 2 のネイティブの単語の発音の認識結果に着目すると，「信頼度 50%以上」の数のなかでも，北米人の A さん，B さん，C さんで差があった．これは，同じ北米英語であっても，出身地域によってイントネーションなどに差があり，多様性が存在することを示している．

表 4 は，音声認識を行った結果の中で，特に認識の信頼度が低かった単語をまとめたものである．これらの結果から，「fan」の「f」などの無声音（声帯振動を伴わない，息だけで出す音），「pen」の「p」などの破裂音が，ネイティブの発音であっても音声認識されにくい音であることが分かる．

表 2　音声認識結果のまとめ 1

		ネイティブ				
		見本音声	Aさん	Bさん	Cさん	Dさん
単語 (26)	第一候補	20	24	21	23	24
	信頼度50%未満	2	3	7	4	6
	信頼度50%以上	18	21	14	19	18
フレーズ (33)	第一候補	32	33	32	33	32
	信頼度50%未満	2	1	1	2	2
	信頼度50%以上	30	32	31	31	30

表 3　音声認識結果のまとめ 2

		日本人	
		Eさん	Fさん
単語 (26)	第一候補	9	10
	信頼度50%未満	7	6
	信頼度50%以上	2	4
フレーズ (33)	第一候補	16	31
	信頼度50%未満	3	1
	信頼度50%以上	13	30

表 4　音声認識結果のまとめ3

confidence	ネイティブ					日本人	
信頼度[%]	見本音声	Aさん	Bさん	Cさん	Dさん	Eさん	Fさん
fan	1.6	91.9	5.2	91.0	2.6	0.8	0.0
hat	4.3	45.4	0.2	80.0	18.1	0.0	0.0
pen	2.6	1.1	1.0	1.6	7.0	0.0	0.0
racket	5.4	36.0	12.8	3.7	45.1	0.0	0.0
box	0.0	0.0	0.0	0.0	76.4	0.0	0.0

5.　利用可能性

本研究の総合結果から，音声認識エンジンを用いた英語の発音評価では，単語については，認識率の低い無声音・破裂音を除いて教材として利用できる．フレーズについては，発音した前後の単語から推測しており，発音が適切でなくても認識してしまっているので，教材としての利用は不適切である．

つまり，ネイティブとノンネイティブの発音において，音声認識の結果に差が出る単語，無声音や破裂音が含む単語を除いた，有声音を中心とする単語に対して利用可能性があることが分かる．

英語の発音指導に対して音声認識エンジンを利用する際に，指導に取り入れる単語の選択が重要となる．北米英語をモデルとするうえで，言語の多様性を認める観点から，ネイティブの発音の中で音声認識の結果で差が出るものは除くべきである．

6.　考察

「Speech Framework」以外の音声認識エンジンも機械学習であるため，本研究と同様の結果になることが予想される．ただし，詳細については調査する必要がある．

7.　おわりに

本研究では，英語の発音指導に対する音声認識エンジンの利用可能性を調査した．ネイティブとノンネイティブの発音において，音声認識の結果に差が出る単語に対して利用可能性があるという研究結果が得られた．小中学校の児童・生徒を被験者とした調査は未検証である．小中学校での教育を通して英語を学んでいる途中の被験者の発音，音声データは非常に有益であるので，調査を進めることで本研究のさらなる追究ができる．

将来的な展望として，「Let's Phonics!!」が持つ録音・再生機能による発音学習に音声認識機能を導入することで機能拡充を行いたい．本アプリケーションをアプリストアに公開することで，より広範囲での利活用を目指すことで，日本の外国語科教育の質的向上に寄与することができる．

参考文献

(1)　文部科学省：“小学校外国語活動・外国語研修ガイドブック”，(2017)．

(2)　平野美沙子：“小学校英語の課題―フォニックスの導入に向けて―”，環境と経営：静岡産業大学論集，22，1，pp.55-66，(2016)．

(3)　一柳佑介：“小学校向けの英語発音学習システムに関する研究～児童へのフィードバック支援機能～”，三重大学大学院工学研究科電気電子工学専攻，令和 2 年度修士論文，(2020)．

(4)　金丸紋子：“プロソディー中心の英語発音指導の効果―日々の授業に取り入れることのできる指導法を求めて―”，日本私学教育研究所紀要，49，pp.25-28，(2013)．

(5)　冬野美晴：“リスニング中心の大学英語科目における英語発音評価ソフトウェアの使用と学生による評価”，西南学院大学言語教育センター紀要，3，pp.21-33，(2013)．

(6)　Apple：“Siri”
https://www.apple.com/jp/siri/（2021 年 7 月参照）

(7)　Amazon：“Alexa”
https://developer.amazon.com/ja-JP/alexa（2021 年 7 月参照）