

# スマートフォン等の音声認識機能を使った発話練習

川 崎 眞理子 (新潟経営大学)<sup>1)</sup>

## 要 旨

本稿では、発話の音声的側面に関する気づきを促す手段として、スマートフォンなどの音声認識機能を活用した取り組みと、学習者の変化と感想について報告する。授業では入力する音声がどのように認識され文字化されるかを指標にして、より正確に認識されるよう練習を行った。また、音声認識では入力音声の強弱は視覚化されないため、入力音声のリズムを確認するために音声合成ソフトウェアの音声波形をも使用した。対象授業は、関西学院大学人間福祉学部で開講している英語コミュニケーションⅠ及びⅡは、外国語選択必修科目で、CEFR-J レベル A1.3程度の学習者を対象に、効果的なコミュニケーションと異文化理解を目標としている。授業では協力的な相手とならば、意思疎通ができる運用能力を身につけるために、ペアやグループで発話練習を行うが、指導者から毎回個別にフィードバックを行うのは残念ながら時間的制約上難しい。そこで、従来の仲間同士でのフィードバックと、適宜の指導者からのフィードバックを組み合わせて、十分な発話練習時間を確保するとともに、音声認識機能を使った練習増量と自分の発話についての気づきを促進することを試みた。その結果、使用教科書中の文章音読の練習回数は増加し、認識成功率は向上した。さらに、アンケート調査から「練習した」や「成功した」など今回の取り組みに対して肯定的な回答が得られた。英語の発音の目標が英語圏の母語話者である必要はないと考えているので、音声認識機能を使って発音を評価するつもりはない。しかし利用できれば便利であるし、手軽で効果的な練習道具になることがわかった。

## 1. 研究の背景と目的

学習者は自分の英語力について、発音ができない、苦手、下手などと述べるのが比較的多い。彼らの基準はほとんど教科書の英語や母語話者の指導者である。果たして、英語学習の目標は母語話者の英語なのか、上手下手の基準は何なのかなど、疑問が生じる。

非母語話者の accent (外国語なまり) については長く論じられてきたが、そのコミュニケーションへの影響についてはまだ研究の歴史は浅いようである。発話の評価要素も、明瞭性 (intelligibility)、comprehensibility (理解性)、あるいは accentedness (アクセント度) などがあり、評価基準も様々である。さらに、発音の要素 (th や r のような個別の音や強弱・高低・連続) のうち、どれがこの評価要素に影響しているのかも報告により異なる。Levis (2005) は

ESLとしては個別音より全体的な音のほうが会話を機能させるためには重要であるが、今日のEIL（国際言語としての英語）としても同様なのかは不明であると主張している。このように目標や評価が明確ではないので、効果的な指導法・学習法など、指導の現場への示唆に乏しく、指導者の多くは直観や経験に基づいて指導している（Derwing & Munro, 2005）。

筆者は英語を使うならば、恐らく国際言語として使うであろう学習者を指導している。しかし、冒頭に記したように目標や評価基準を「ネイティブ」とする学習者が多い。毎年毎学期、新しい学習者との出会いの時に、この点について、話し合うこともしばしばである。Chong (2014)が記すように、2億人が英語を使う中、ほとんどが非母語話者である時代に、母語話者の英語に到達する必要はない。Jenkins (2002) がずいぶん前に主張したとおり、筆者も様々な発音があっいていいし、そのような多様な英語に触れていなければならないと考えていると伝えたくて、「しかし、とにかくコミュニケーションを成立させなければならない」ことを確認する。

すなわち、本取り組みの対象である英語コミュニケーションという科目を履修したならば、英語でコミュニケーションを円滑に行う力を養ってもらいたい。そのためには、理解してもらいやすい発音をすることも目標の一つである。理解してもらいやすい発音はどんな発音かを理解する必要がある。そこで授業内外での活動の一部として、音声認識機能や音声編集ソフトウェアを使って練習し、学習者の発音に対する気づきとわかりやすい発音への改善を目指した。

## 2. 対象科目

対象科目、英語コミュニケーションⅠ及びⅡは、人間福祉学部1年次の外国語選択必修科目である。本科目の他に、スペイン語、韓国語、手話等がある。2年次には引き続き英語コミュニケーションⅢとⅣを履修する。履修条件としてⅠ、Ⅱ、Ⅲ、Ⅳの順で単位を取得しなければならない。なお、募集は3クラスで、抽選により履修者を決めている。

### 2.1 クラス編成

今回は筆者が担当する英語コミュニケーションⅠ及びⅡの2組と3組を対象とした。履修確定者を入学時のGTEC（リスニング・リーディング）スコアにより上位を1組、下位を3組、中間を2組としそれぞれ約25名前後になるように分けている。ただし、1組は例年留学希望者や言語教育センター開講科目履修のために英語コミュニケーションⅡを受講しない者がいるので人数は若干多く、逆に3組は初級レベルの学習者が混在するので、指導の便宜上若干少なくしている。2組と3組は、実際の授業運営においては、いずれも初中級とした。

### 2.2 授業目的と目標及び授業形態

2組と3組の春学期の英語コミュニケーションⅠと、秋学期の同Ⅱの到達目標はそれぞれCEFR-JのA2.1とA2.2である。当該科目の目的と到達目標は次のとおりである（2018年度シラバスより抜粋）。

### 英語コミュニケーション I (春学期)

#### 授業目的：

本授業の目的は、複雑でない状況において、必要に応じて学習した異文化知識を適用して、適切にやりとりができるコミュニケーションスキルを身につけることである。

#### 到達目標：

- なじみのあるトピックならば、ある程度楽にフレーズを組み立てて、明らかに長い言いよどみや言いなおしはあるものの、短いやりとりができる。
- 短い社会的やりとりを日常レベルの礼儀正しい挨拶や呼称を使ってできる。
- 招待の授受や、提案、謝罪等ができる。
- 協力的な相手と、簡単な情報交換をし、異文化について知ることができる。

### 英語コミュニケーション II (秋学期)

#### 授業目的：

本授業の目的は、さまざまな状況において、必要に応じて学習した異文化知識を適用して、適切にやりとりができるコミュニケーションスキルを身につけることである。

#### 到達目標：

- 沈黙や言いなおしが多いながらも、短い発話を理解してもらうことができる。
- 最も簡単な一般的な表現を使い、基本的規則に従って、簡単にしかし効果的に社交ができる。
- 言葉を使って基本的なこと、例えば情報交換したり、お願いをしたり、意見を言ったりできる。
- 協力的な相手と情報を交換したり意見を言ったりし、異文化について知識を得ることができる。

#### 授業形態：

すでに持っている語彙を含めた英語の知識を最大限活用して、最低限理解してもらえる発話をするを第一の目標とする。加えて、コミュニケーションにおいては文化的違いを理解しておくことが不可欠であるから、英語で多文化に関する知識を学習する。使用教科書は、People Like Us (ELT 出版) で、様々な国の母語・非母語話者が特定の文化的テーマについて話しており、それをもとに使用表現を学習し、テーマに沿って意見交換を行った。時には時間をかけて調査をし、結果を発表し合った。常に、指導者や教科書 CD の音声はもとより、クラスメイトの発話にも注意を傾け、自分の発話の改善に努めるよう促した。自分の発話やグループ活動中のやりとりは適宜録音し、発話に対して包括的な評価とフィードバックを行った。

### 3. 音声認識機能の活用

今回、授業改善の取り組みとして、スマートフォンの音声認識機能を使って発話を視覚化することで、練習への取り組み姿勢や発話が改善されるかどうかを追跡した。だれもが負担なく簡単に活用できるよう、全員が所有していたスマートフォンの音声認識機能を利用して新しいメールに英文を音声入力してみた。iOS と Android が混在していたが、ほぼ同様に使えた。音声認識機能については約半数が程度の差はあるものの知ってはいたが、実際に日本語でも使用経験があるものは少なく、使用方法の説明が必要であった。特に、英語モードに設定してから入力するよう注意が必要であった。また、当該機能の特性として、語列から予測して変換されるようで、語

単位での入力より句、節、さらには文単位での入力のほうが正しく認識されやすいことにも言及した。したがって、途中で言い誤った場合は、意味の切れ目までさかのぼって再入力するよう指示した。あとは、練習しながら、どのような入力が認識されやすいのか、学習者の自主的な試行に任せた。いろいろな発見があったようである。この間、練習回数は増えていた。しかし、少数ではあったがどうしても正しく認識されない事例があった。他の機材を使用しても同様の結果であり、理由は不明で最後まで解決されることはなかった。

音声認識機能の利用に加えて、入力音声の強弱や無音時間（ポーズ）を視覚化し、モデル音声と比較することで自らの音声を客観的に評価できるように、音声編集アプリを使用する計画であった。しかし、全員がアプリをダウンロードする必要があり、これを強制することの是非が不明であったし、万一の不具合発生時に対応できないため、今回は、音声入力中に画面下部に表示される音声検出表示に注目させるにとどめた。PCに Audacity という無料音声編集ソフトウェアを紹介し、任意で各自のPCにインストールして授業外で練習することを勧めた。

## 4. 調査方法

### 4.1 音声認識用素材と測定方法

調査は音読音声の評価とし、授業に使用した教科書から調査時期に学習していた部分で説明文を抜粋して使用した。難易度や音声的側面について一切調整を行っていないが、選択時点で、固有名詞が含まれないよう留意した。それぞれ次のとおりである。

#### <春学期 英語コミュニケーションⅠ>

Women who are not working go shopping every day. It's a social occasion with the separate trips to the bakery, supermarket, bank, or butcher. (24語)

試訳：働いていない女性は毎日買い物に行く。これは彼女らにとって社会にでる機会、ベーカリーやスーパーマーケット、銀行、精肉店などに個別に出かける。

#### <秋学期 英語コミュニケーションⅡ>

Schools, cities and other organizations sponsor team sports for young people and professionals. Sports are an important part of the culture. In general, most Americans spend a large amount of their leisure time socializing or watching television. (37語)

試訳：学校、市、その他の組織は若者やプロのスポーツチームのスポンサーになる。スポーツは文化の重要な一部である。一般にほとんどのアメリカ人は余暇の大部分を社交したりテレビを見たりして過ごす。

正しく読めているかを確認し、その後数回練習し、その場でメールの本文に音声認識による入力を行い、筆者に送信してもらった。春学期は入力時に隣席の者が iPhone で音声を録音し、その音声ファイルも筆者に添付送信してもらった。全員の入力結果を集約して印刷したものを配布し、誤認識の原因として考えられることを伝えた。秋学期は時間の都合上、録音の提出は求めなかった。

## 4.2 年度末アンケート調査

秋学期最終授業で、次の5つの質問の回答（5択）と、その他感想があれば記述するように依頼した。

### 質問1

音声認識機能を知っていましたか。

- 1 知らなかった、2 あまり知らなかった、3 聞いたことはあった、4 知っていた、  
5 よく知っていた

### 質問2

音声認識機能の使い方を知っていましたか。

- 1 知らなかった、2 あまり知らなかった、3 聞いたことはあった、4 知っていた、  
5 よく知っていた

### 質問3

音声認識機能に自分の英語を書き起こしはうまくいきましたか。

- 1 いかなかった、2 ややいかなかった、3 どちらでもない、4 ややいった、5 いった

### 質問4

音声認識機能を使って練習をしましたか。

- 1 しなかった、2 あまりしなかった、3 どちらでもない、4 した、5 かなりした

### 質問5

音声認識機能を使ったことで音声面が上達したと思いますか。

- 1 しなかった、2 あまりしなかった、3 どちらでもない、4 した、5 かなりした

## 5. 結果と考察

### 5.1 音声認識測定

英語コミュニケーションⅠ及びⅡで春秋両データがある者は2、3組合計  $N_1=35$ （男26女9）であった。協力者別成功率及び単語別成功率の記述統計値はそれぞれ表1と2のとおりである。同じ素材を使っていないので比較はできないが、秋学期の平均音声認識成功率が高く（春0.67秋0.78）、最低成功率も高い（春0.29秋0.49）。しかし、ばらつき（標準偏差0.16）は同じである。単語別の平均成功率も秋学期のほうが高い（春0.67秋0.78）。どうしても納得いかず、やりなおさせてほしいと申し出る者が多かった。公平な評価のためには、測定方法を統一すべきであるが、授業としての実施であり、学習者の意思を尊重し、納得いくものを提出してよいことにした。どうしても認識されない声の持ち主もおり、端末の交換や、音量調整を試みたが、現時点でも原因は不明のままである。

春学期の各語の認識成功率を図1に示す。また、下の文中で囲いをした語は誤りが多かった内容語ある。成功率は occasion、butcher とともに、0.29であった。

Women who are not working go shopping every day. It's a social occasion with the separate trips to the bakery, supermarket, bank, or butcher.

表1. 協力者別音声認識成功率 (N<sub>1</sub> = 35)

	春学期 (英語コミュニケーション I)	秋学期 (英語コミュニケーション II)
平均	0.67	0.78
最高	0.83	0.97
最低	0.29	0.49
標準偏差	0.16	0.16

表2. 単語別音声認識成功率

	春学期 (英語コミュニケーション I)	秋学期 (英語コミュニケーション II)
総語数 N <sub>2</sub>	24	37
平均	0.67	0.78
最高	0.97	1.00
最低	0.17 (内容語0.29、機能語0.17)	0.11 (内容語0.11、機能語0.31)
標準偏差	0.23	0.20

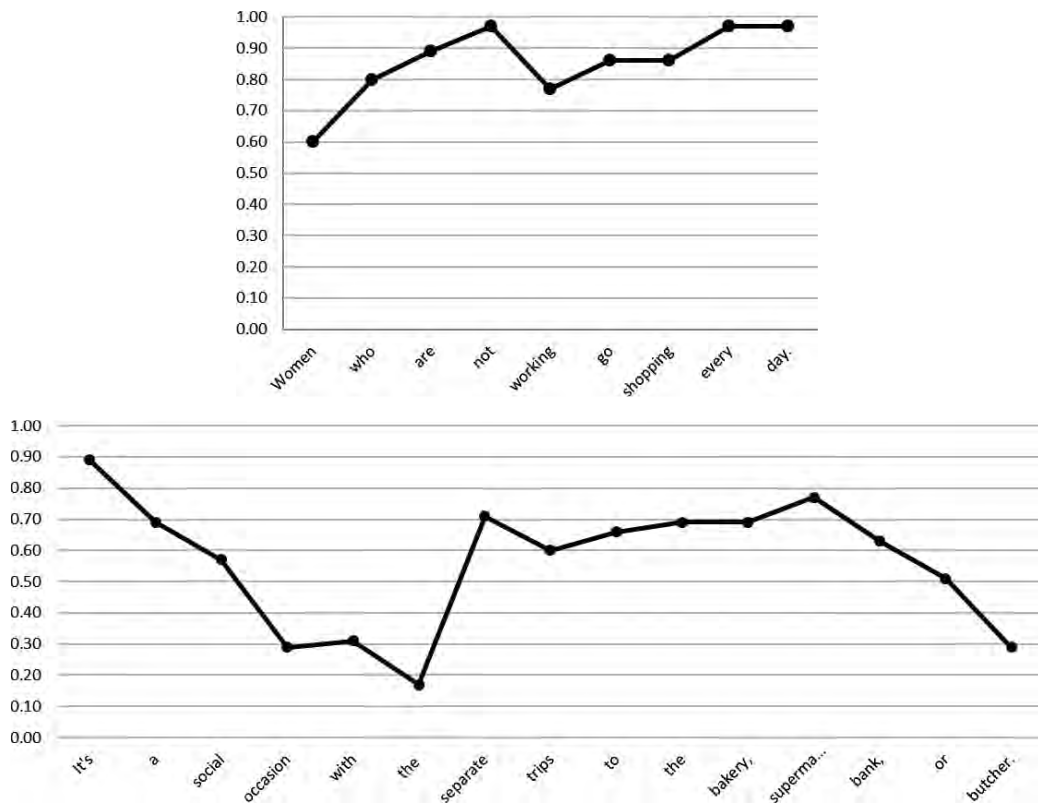


図1 単語別音声認識成功率：春学期 (英語コミュニケーション I)



成功率が低かった occasion (0.29) は location (0.17) や OK~ (0.31) と書き起こされていたが、前者は先行する social の最後子音を soft L ではなく hard L と発音したため次の o とともに lo と書き起こされたと考えられるものもあったが、social が入力されていないものもあった。一方、location が OK song や OK soon となったり、butcher が picture と誤認識されたりする原因は、頭子音の弱さや母音の発音であろう。成功率が低くはなかったが、同様の誤りが見られたのは、bank で、funk, bunk, punk, drunk と認識されていた。

この3語ほど成功率は低くはないが、内容語としては social (0.59) が次に低かった。機能語としては with the の成功率が低く (0.31と0.17)、the が成功した者は1名を除き with も成功し、with に成功した者の半数は the が入力されていなかった (録音確認)。

秋学期の各語の認識率を図2に示す。また、下の文中で、下線を引いた語は成功とみなしたが、形態素の誤りが見られた語である。

Schools, cities and other organizations sponsor team sports for young people and professionals. Sports are an important part of the culture. In general, most Americans spend a large amount of their leisure time socializing or watching television.

Schools は school が0.23と最も多かったが、成功として扱った。その他の誤りは、scooter, scooters, score, scores であった。同様に sponsor は sponsor a (0.34) や sponsored (0.26) が多く、成功に含まれている。さらに young は0.31が younger であった。いずれも語尾に母音を付加したことが原因と考えられる。そのほか an important の an (0.11) は入力されていなかった。television は TV と認識されていたが、恐らく、入力時に TV と発音したのであろう。

特殊な例として、一語の誤認識が続く何語かの連続的な誤りを招いたと思われる事例があった。音声認識機能が、入力に基づき予測変換をするためと思われる。全体的には明瞭な、強弱がはっきりした発音が正しく認識されやすいようであった。

## 5.2 年度末アンケート調査結果並びに自由記述回答

音声認識機能を知っていたのは2組では75%、3組では35%にとどまっていた。さらに使い方になると、「だいたい知っていた」から「よく知っていた」まで合わせても2組で40%、3組では30%であった。成果については、2組3組ともに認識がうまくいったとの回答が約80%であった。この機能を使って練習したかの問いには、2組90%近くが練習したとした半面、3組では約50%であったが、しなかったとの回答は両組とも10%以下であった。自由記述により、次のような感想が寄せられた。

- 今まで知らなかった英語の勉強方法を知ることができた。
- 音声機能を使ったことで自分の発音を確認できた。
- 音声認識機能やリスニング・スピーキングテストで自分の英語力の低さがわかったので、特に話す、聞く能力を養いたいと思います。

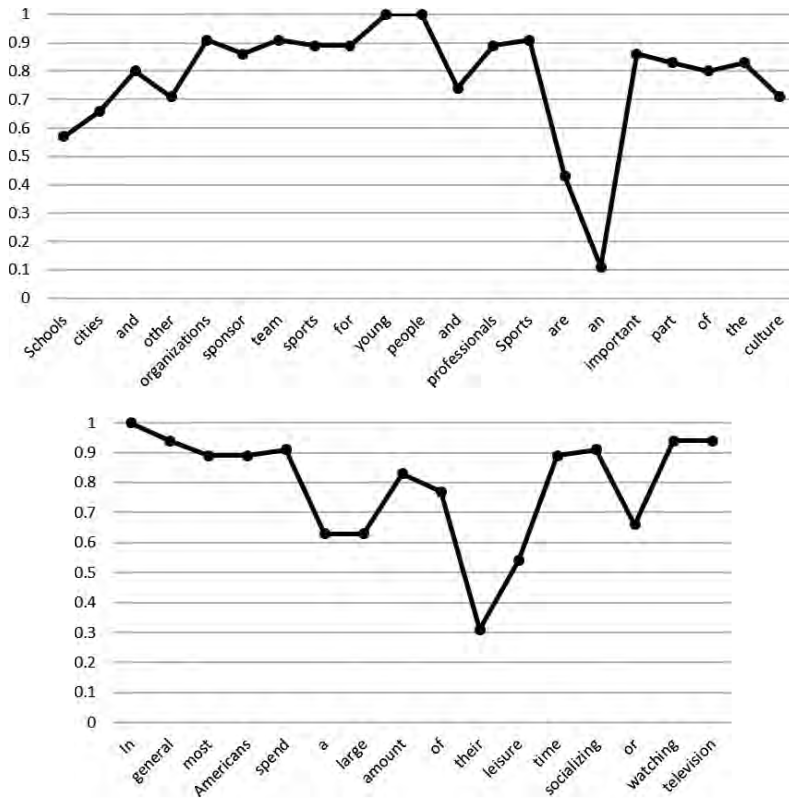


図2. 単語別音声認識成功率：秋学期（英語コミュニケーションⅡ）

アンケートと自由記述両方から、音声認識機能を使うことに肯定的な結果であった。目新しいさかもしれないが、反復練習の効果はあったようである。

## 6. まとめ、制約および今後の課題

本研究では、音声認識アプリを利用して音読の文字化を試みることで、英語の音への意識とその改善を目指した。その結果、学習者に練習を促すことができ、音声認識の成功率の改善も見られた。何より、自主的に楽しく練習している様子を見ることができた。

しかし、授業の改善が目的で、授業計画内での練習と測定とし、使用教科書内の文を測定素材と使用したため、実証研究として音声認識機能活用の効果を論じることはできない。正確に発音の変化や改善を測定できてはいない。加えて、音声認識機能のアルゴリズムは開発者のみを知るものであり、例えば、どの程度の精度で認識し文字化するのか、誤った文字化の原因は何かなどが不明で、その性能／精度については批判が多い。一例としてドイツ語の発話評価ソフトウェアについて、ドイツ語母語話者の評価が低いこともあるとの報告があり (Wildner, 2002)、英語のソフトウェアでは上級の話者に対する評価に Mackey & Choi (1998) が疑問を呈している。本取り組みにおいても、正しく文字化されなかったからといって、学習者の発音に問題があるというわけではなかった。筆者自身が音声聞いた限りでは、コミュニケーションに支障がないと思われる発音でも正しく文字化されない事例が多々あった。



最後に、研究の背景で記述したとおり、第二言語や国際語として英語を使用している人たちが多い今日、母語話者の発音に準じる必要はないと考えている。一方で、音声認識機能を活用できれば便利で需要は高まるであろうから、この機能を活用するための発音要件や、その発音はコミュニケーション能力の評価に使えるかなど、明らかになってほしい。機能の改良進歩を見守りつつ、指導の現場で活用し、学習者の利益につなげていくと同時に、実証研究も考えていきたい。

## 謝辞

本授業改善の実践にあたり、関西学院大学高等教育推進センターの2018年度先端的な授業改善実践研究助成をいただきましたことを感謝申し上げます。

## 注

- 1) 2018年度は関西学院大学人間福祉学部助教

## 参考文献

- Chong, C-S. (2014). Whose accent is better? *English Teaching Professional*, May, 2014. Retrieved from: <https://www.etprofessional.com/whose-accent-is-better>
- Derwing, T. M. & Munro, M J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379-398.
- Jenkins, J. (2002). Global English and the teaching of pronunciation. TeachingEnglish, Retrieved from: <https://www.teachingenglish.org.uk/article/global-english-teaching-pronunciation>
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369-377.
- Mackey, A. & Choi, J.-Y. (1998). Review of TriplePlayPlus! English, *Language Learning & Technology* 2, 19-21.
- Wildner, S. (2002). Learn German Now! Version 8, Software review, *CALICO Journal*, 20, 161-174.