

令和 2 年 6 月 9 日現在

機関番号：13901

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K02881

研究課題名(和文) シャドーイングにおけるスマホの音声合成・音声認識機能の利用効果に関する実証的研究

研究課題名(英文) An Empirical Study on the Effectiveness of Using Smartphone Speech Technologies in EFL Shadowing Activities

研究代表者

古泉 隆 (Koizumi, Takashi)

名古屋大学・教養教育院・講師

研究者番号：60549541

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：音声合成・音声認識機能を外国語シャドーイング活動にどのように取り入れることができるか、またそれらを利用したシャドーイング活動により、どのような学習効果が得られるかを実証的に検証した。第一に、音声認識を利用した独自のシャドーイング自動評価を検討し、その自動評価と音声合成を活用したシャドーイング学習アプリを開発した。次に、大学生にアプリを使ってシャドーイング活動をしてもらった結果、音声合成でも母語話者音声と同程度スピーキング力向上などの効果が得られること、さらに、音声認識を用いた自動評価を活用することで、スピーキング力や動機付けという点でより効果が期待できる可能性が示唆された。

研究成果の学術的意義や社会的意義

近年、iPhoneやGoogleの音声検索にみられるように音声認識・音声合成技術は急速に進歩し実用化されている。今後も進歩が期待される音声技術を、外国語学習に活用する試みはすでにみられるが、実証的に検証した例は少ない。本研究では、シャドーイング活動という外国語学習法に、音声技術をどのように利用できるか、また、どのような点で効果があるのかを実証的に検証した。得られた結果は、今後の語学教材開発・語学学習において活用が可能である。

研究成果の概要(英文)：This study examined how text-to-speech and speech-to-text technologies can be incorporated into EFL shadowing activities and how effective they are for language learning. Using cloud-based speech technologies, we developed a smartphone app that can evaluate learners' shadowing performances automatically. We then asked university students to practice shadowing using the app to determine the effectiveness of their learning. The results suggested that text-to-speech voices were as effective as native speakers' voices for improving learners' speaking abilities and that automatic evaluations using speech-to-text technologies were more effective than self-evaluations for both improving their speaking abilities and motivating students.

研究分野：英語教育

キーワード：シャドーイング 音声認識 音声合成

1. 研究開始当初の背景

(1) スマートフォンの普及が急速に進むと同時に、iPhone や iPad の iOS に搭載された Siri や Google の検索においても音声認識・音声合成技術を用いたサービスが実用化され始めた。

(2) シャドーイングは音声活動を中心に行う語学学習法であり、日本の英語教育においてすでに広く利用・認知されていた。その中で、音声合成や音声認識技術をシャドーイング活動・教材開発に利用した事例や報告は少数のみであった。

2. 研究の目的

本研究は、音声合成・音声認識機能をシャドーイング活動にどのように取り入れることができるかを検討し、それらを利用したスマホでのシャドーイング活動により、どのような学習効果が得られるかを実証的に検証することを目的とする。

3. 研究の方法

(1) 研究方法としては、まず、音声認識をどのようにシャドーイング活動に導入できるかを調査した。大学生のシャドーイングの録音音声に対して、音声認識による評価と教員による評価とを比較し、音声認識を利用することでシャドーイング評価をどの程度適切に行えるか調査した。

(2) 次に、音声認識を利用してシャドーイングの出来具合を自動評価する仕組みを開発し、音声合成によるお手本音声とともに、それらを活用したスマホアプリを開発した。従来型(母語話者のお手本を聞いてのシャドーイング+録音によるシャドーイング自己評価)、音声技術利用型(音声合成のお手本を聞いてのシャドーイング+音声認識機能によるシャドーイング自動評価)、折衷型(音声合成のお手本を聞いてのシャドーイング+録音によるシャドーイング自己評価)の3タイプのアプリを開発した。

(3) 大学生(3グループ)に計10日間、開発したアプリでシャドーイング活動をしてもらい、事前・事後にリスニング・スピーキング・コロケーションのテストおよび事後アンケート・インタビューを行った。その後、実験結果を分析した。

4. 研究成果

(1) 音声認識によるシャドーイング自動評価の検討(古泉, 2018)

研究開始当初はスマホ(ブラウザ)の音声認識APIを利用する予定であったがAPIの仕様変更により意図したように動作しないことが判明した。そのため、より安定的かつ、汎用的にスマホおよびPCでも利用可能なクラウド型の音声認識サービスを利用することとした。本研究では、代表的なクラウド型音声認識サービスである Google Speech Cloud、IMB Watson Speech To Text、Microsoft Azure Bing Speech の3つを比較検証した。

日本人大学生による英文音読およびシャドーイングの録音データに対して、各音声認識サービスで音声認識を実施し「認識率」を算出した。ここでの「認識率」とは、正しく認識された単語数をシャドーイングテキストの単語数で割った数値とした。音読データに対する認識率は、Google が約71%、Watson が約57%、Azure が約65%であった。さらにGoogleでは、単語やフレーズを「ヒント」として与えることで音声認識の精度を向上させる機能がある。この「ヒント」を用いることで認識率が約81%となった。

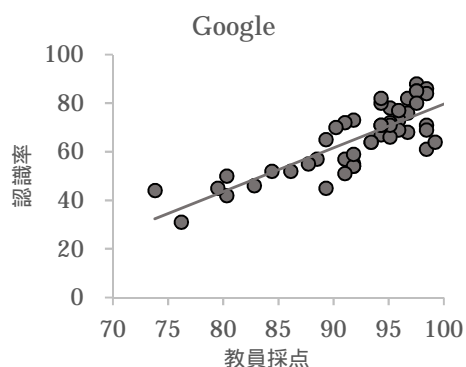


図1: 教員による採点と音声認識率の相関図

次に、英語教員によるシャドーイングの採点と音声認識による認識率との相関を調査した。英語教員による採点は音節法にならない、シャドーイングテキストの単語を音節ごとに分割し、音節単位で行った。Googleでは相関係数(ピアソンの積率相関係数)が $r = .82$ を示し(図1)、Google ヒント有においても $r = .80$ となり、かなり強い相関が見られた。Watsonでも、 $r = .73$ と比較的強い相関が見られた。一方で、Azureでは、 $r = .20$ と弱い相関しか確認されなかった。以上の結果から、Googleといった音声認識サービスを利用することで、教員が音節法で採点した場合と強い相関のある評価が可能であることが示された。

(2) シャドーイングアプリの開発

続いて、Google Speech Cloudの音声認識サービスを利用した自動評価を組み込んだAndroidスマホ用シャドーイングアプリを開発した。コロケーションを組み入れたシャドーイング英文(各およそ100語)を20個用意した。また、シャドーイングタスクとして、1)リスニング、2)英文を見ないシャドーイング、3)シャドーイング録音(シャドーイングの出来を評価)、4)英文・訳の確認、5)英文を見て意味を意識してシャドーイング、6)英文を見ないで意味を意識してシャドーイング、7)シャドーイング録音(シャドーイングの出来を評価)の7つの

ステージを設けた。各ステージを順に進んでいき、ステージ7を終えると一つのユニットが完了する(全20ユニット)。ステージ3と7の評価においては、学習者が自分の録音を聞いて自己採点する方式(一定語数ごとの単語が言えているかをチェックしていく)と、音声認識を利用した自動評価方式の2つのバージョンを作成した。自動評価方式では、シャドーイングのスコアを表示するとともに、言えていなかった(音声認識できなかった)箇所をマーク表示し、文全体またはマーク箇所の発音を音声合成で聞いて練習できるようにした。シャドーイングのお手本音声に関しても、母語話者による録音と合成音声の2つのバージョンを用意した。音声合成にはクラウド型サービスのAmazon Pollyを利用した。最終的に、1)従来型(母語話者音声+シャドーイング自己評価)、2)音声技術利用型(音声合成+シャドーイング自動評価)、3)折衷型(音声合成+シャドーイング自己評価)の3つのモードが利用できるアプリを開発した。実験の際には、ログインIDを配布し、IDにより決まった型(従来型・音声技術利用型・折衷型)の学習となるようにした。

(3) 音声合成・音声認識の利用効果

日本人大学生(3グループ)に、開発したアプリを使って計10日間シャドーイング学習を自宅等で行ってもらった。1日に2つの英文(ユニットを2つ)のシャドーイングを行ってもらった。スマホは事前に準備し、同型のものを実験参加者に貸し出した。従来型グループ、音声技術利用型グループ、および折衷型グループの3グループで学習を行ってもらった。事前・事後テストとして、リスニング能力測定にCASEC、スピーキング能力測定にVersant Speaking Testを実施した。また、筆記によるコロケーションテストおよびアンケート・インタビューを事後に行った。

結果として、リスニング能力については、従来型、音声技術利用型、折衷型ともに、事前・事後の得点において統計的に有意な伸びはなかった。一方、スピーキング能力については、4つのサブスキル(Sentence Mastery, Fluency, Vocabulary, Pronunciation)のうち、Sentence Mastery(文章構文)およびFluency(流暢性)において事前・事後で有意な伸びがあったが、グループの違いと得点伸びの間には有意な交互作用は見られなかった。すなわち、Sentence MasteryおよびFluencyにおいては、シャドーイング活動によって得点が伸びたが、グループの違い(音声合成・音声認識の利用の有無)によって有意な差がなかったといえる。しかし、Sentence Masteryにおいては、効果の大きさを示す効果量を比較すると音声技術利用型の効果量が大きいことがわかった。音声技術利用型と折衷型の違いは、ステージ3と7の評価方法・タスクである。すなわち、Sentence Masteryというスピーキングスキルにおいては、音声認識による自動評価および付随する音声認識・音声合成を活用した練習タスクが、一定程度効果が期待できる可能性が示唆された。その他のスピーキングのサブスキルであるVocabulary(語彙)およびPronunciation(発音)においては事前・事後で有意な伸びはなかった。次にコロケーションテストに関してであるが、事後テスト(日本語に対応したコロケーションを書いてもらう)の結果、3つのグループ間で正解率に有意な差はなかった。

アンケート・インタビュー分析からも、いくつか特徴的なことがわかった。「シャドーイングのお手本の音声の速度を自分で変更できる機能は必要だと思いますか」という問いに対して、「是非必要」「必要」が65%で、不要が35%であった。本研究で開発したシャドーイングアプリには音声速度を自分で変更できる機能は備わっていないが、必要と感じている学生が比較的多い結果となった。一方で、少なからずの学生は不要と感じており、その理由として「自分で変更できるとゆっくりにしてしまうので、強制的にスピードがあったほうがよい」などの意見がみられた。技術的には、MP3などの録音ファイルで速度調整するよりも、音声合成を用いることでポーズ時間を調整するなど細かい変更ができ、より自然な速度調整が可能である。リスニング練習の際には、再生速度を学習者自身がコントロールすることで理解度が向上することが報告されているが(Zhao, 1997)、シャドーイング活動において、お手本音声の速度調整を自分で出来ることが利点になるのかについて、今後の研究課題として検証していきたい。次に、「(シャドーイング活動において)自己評価・自動評価する機会はあったほうがよいですか」という問いに対しては、ほとんどの参加者が「是非あったほうがよい」「あったほうがよい」と肯定的な回答をした。理由を分析すると、各グループともに「シャドーイングでどこが出来ていなかったのかわかる」など、自分の現状や弱点がわかる点をあげている回答が共通して多く見られる一方で、自動評価を行った音声技術利用型のグループでは「やる気がでる」「達成感が得られる」など、学習意欲や動機付けにつながるなどの回答が他のグループより顕著に多かった。リスニング・スピーキングの事前・事後テスト結果からは、シャドーイングを自己評価するグループと自動評価するグループとでは、得点の伸びに有意な差は見られなかったが、意欲・動機付けという心理的な面で見ると、自動評価を用いることは有効に働いていると考えられる。

(4) まとめと今後の展望

本研究では音声認識を利用した独自のシャドーイング自動評価を検討し、その自動評価と音

声合成を組み込んだシャドーイング学習アプリを開発した。大学生にアプリを使ってシャドーイング活動をしてもらった結果、音声合成によるお手本音声でも母語話者音声のお手本と同程度、スピーキング力向上などの効果が得られること、さらに、音声認識を用いた自動評価を活用することで、スピーキング力や動機付けという点でより効果が期待できる可能性が示唆された。また、アンケート調査の結果、シャドーイング活動においてお手本音声の速度調整を自分で出来ることが利点になるのか否かという新たな研究視点が得られた。お手本の速度を変えるという点については、学習者の手動によるものだけでなく、音声認識による自動評価の結果に基づいて自動的に変更することも可能である。今後の展開として、お手本の速度の変更の有無、および自動評価を活用するなどの変更の手段に着目し、さらに音声技術の利用の可能性・方法を調査していきたい。

< 引用文献 >

古泉 隆、「シャドーイング自動評価システムの開発に向けたクラウド型音声認識サービスの精度および特性の比較検証」、『外国語教育メディア学会(LET)中部支部研究紀要』、2018、11-24

Zhao, Y., The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics*, 18(1), 1997, 49-68.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 古泉隆	4. 巻 29
2. 論文標題 シャドーイング自動評価システムの開発に向けたクラウド型音声認識サービスの精度および特性の比較検証	5. 発行年 2018年
3. 雑誌名 外国語教育メディア学会(LET)中部支部研究紀要	6. 最初と最後の頁 11-24
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 古泉隆
2. 発表標題 シャドーイング活動における偶発的コロケーション学習の検証
3. 学会等名 LET第92回（2018年度秋季）中部支部研究大会
4. 発表年 2018年

1. 発表者名 古泉隆
2. 発表標題 スマホによる音声認識の精度検証と英語学習への利用の検討
3. 学会等名 LET中部支部第89回（2017年度春季）支部研究大会
4. 発表年 2017年

1. 発表者名 古泉隆
2. 発表標題 音声合成・音声認識の利用効果を検証するためのスマホアプリ開発
3. 学会等名 LET中部支部第90回（2017年度秋季）支部研究大会
4. 発表年 2018年

1. 発表者名 古泉隆, 石田知美, 松原緑, 小松雅宏, 杉浦正利
2. 発表標題 アカデミックプレゼンテーション表現学習のためのスマホ対応ウェブアプリ教材の開発
3. 学会等名 外国語教育メディア学会 (LET) 第56回全国研究大会
4. 発表年 2016年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
連携研究者	杉浦 正利 (Sugiura Masatoshi) (80216308)	名古屋大学・人文学研究科・教授 (13901)	