

令和 5 年 6 月 7 日現在

機関番号：17102

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19936

研究課題名(和文) 英語母語話者の調音と音象徴を伝達するVR技術を基盤とした音声訓練法の設計

研究課題名(英文) Design of a speech training method using VR technology for transferring articulation and sound symbols of native English speakers

研究代表者

福島 政期 (Fukushima, Shogo)

九州大学・システム情報科学研究院・准教授

研究者番号：30761861

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：英語学習者が正しい調音や言語感覚を学ぶために、英語母語話者の音声や顔を学習者のものに変換して提示する調音学習システムを検討した。また、英語の言語感覚を理解するために、バーチャル空間で身体や道具を使うことで単語の意味を再現しながら学習するシステムを検討した。前者は、英語母語話者の音声を学習者の音声に変換する技術開発およびシャドーイング学習への適用可能性について検証した。後者は英語の言語感覚を理解するために、バーチャル空間で身体や道具を使うことで単語の意味を再現しながら学習するシステムを実装・評価するとともに、バーチャル環境のオブジェクトから学習可能な他動詞を生成することに取り組んだ。

研究成果の学術的意義や社会的意義

本研究では特にバーチャルリアリティ技術を使って英語の学習手法をデザインし直すことに取り組み、その学習効果として記憶の再生テストを使った評価を行った。今後HMDなどのバーチャルリアリティ技術が日常のあらゆる場面で利用されることを想定し、どのような手法で人の学習支援に利用できるかを実機を使いながら評価したことが本研究の学術的貢献である。

研究成果の概要(英文)：In order for English learners to learn correct articulation and language sense, we examined an articulation learning system that transforms and presents the native English speaker's voice and face into that of the learner. We also studied a learning system that reproduces the meaning of words by using the body and tools in a virtual space in order to understand English language senses. In the former case, we developed a technology to convert a native English speaker's voice into a learner's voice, and verified its applicability to shadowing learning. For the latter, we implemented and evaluated a learning system that reproduces the meaning of words in a virtual space by using body and tools in order to understand the sense of English language, and also worked on generating learnable transitive verbs from objects in the virtual environment.

研究分野：バーチャルリアリティ

キーワード：バーチャルリアリティ 学習支援 第二言語習得 語学学習

1. 研究開始当初の背景

日本人の TOEFL のスコアは、アジア 28 カ国中最下位で、スピーキングの得点は世界 147 カ国中最下位であるという報告がある[松香, 2010]。日本語は英語との言語間距離が遠い。中学校と高校で 875 時間（中学：50 分×週 4 日×35 週×3 年、高校：50 分×週 6 日×35 週×3 年）を学んだとしても、その時間は極めて短い。そのため当然、授業での音声練習の機会も少なくなってしまう。

自主学习での音声練習の場合、教師が正しい調音に導くことが難しいため、どうしても自分なりの発音になってしまう。その結果、英語母語話者の発音と自身の調音が一致せず、音声の聴取や発話が困難になる。学習者の発音の正誤を判定する音声認識システムを利用することも考えられるが、正しい調音に導くことは難しい。発音学習用の教師の音声と比較する研究では、教師の性別や音声ピッチ、発話スピードによって学習効果が変わることが報告されている[Wang, 2010]。また、学習者の音声や発話速度に類似した教師は Golden speaker と呼ばれ、発音学習の効率が良いとされている[Probst, 2002]。本研究ではこの Golden speaker を発展させ、学習者自身を教師とすることで調音を形成しやすくなるのではないかと考えた。

英語の上級者は母語を介さず、英語とその言語感覚などで思考し、英語を出力する[potter, 1984]。また英語に習熟した学習者ほど、脳の言語野だけでなく、体性感覚野などの他の感覚野が活性化する[有路, 2015]。つまり英語を“知っている”状態から“使える”知識に変換するためには、英語とその訳の対を暗記するだけでなく、英語を体験や身体動作と共に記憶することが重要であると考えられる。

2. 研究の目的

本研究は、英語を第二言語とする学習者の学習支援を目的とする。そのために、音声練習において、英語母語話者の声や顔を学習者のものに変換し、学習者にとっての正しい調音を直接提示する（研究項目 1）。また、単語学習において、バーチャル空間で身体動作を使いながら暗記する手法を実現する（研究項目 2）。

3. 研究の方法

<研究項目 1: 正しい調音を伝える声と顔の変換技術>

まずは英語母語話者の音声や顔を学習者のものに変換して提示する技術を実現する。本技術については音声変換を専門とする名古屋大学の戸田教授と共に開発する。統計的音声変換処理を活用して音声を変換する。まずは、入力話者と出力話者が同一言語で発声した音声対で実装し、その後、深層学習に基づく変分自己符号化器を改良することで、英語母語話者の英語音声と英語学習者の母語音声をを用いた学習処理を実現する。また、体表密着型マイクロフォンを用いて英語学習者の体内伝導音声の収録を行い、英語母語話者から英語学習者の体内伝導音声への変換処理も実現する。

次に変換技術の学習効果を評価する。本変換手法で学習するグループと英語母語話者のもので学習するグループで学習効果を比較する。学習前の発話と学習後の発話を収録し、発音判定ソフトウェア (GlobalvoiceCALL2) および英語母語話者の得点で比較する。音素・アクセント・イントネーション・タイミングの 4 側面が学習の前後で向上することを確認する。

<研究項目 2: バーチャル空間での実演による学習手法>

語句の意味を、動作を使いながら学習する手法に実演効果がある。例えば”Raise the arm”という語句を覚える際に実際に手を上げる動作をしながら覚えるものや、”Look in the Mirror”のような道具を使いながら覚えるものなどがある。実演効果は単に語句を口頭で繰り返す手法よりも記憶の再生テストのスコアが向上することが示されている。本研究では、この実演をバーチャル空間で行うことで、実世界では体験することが難しい他動詞を学習できる手法を実現する。また、バーチャル空間における実演効果において学習可能な他動詞を機械的に生成するための手法も検討する。

本手法の学習効果を評価するために、学習直後の再生テスト（英単語をみて、日本語訳を回答するテスト）に加えて 1 週間後にも同様の再生テストを行い、本手法が記憶保持に与える影響を確認する。

4. 研究成果

<研究項目 1: 正しい調音を伝える声と顔の変換技術>

英語母語話者の音声を学習者の音声に変換する技術開発およびシャドーイング学習への適用可能性について検証した。音声変換は話者と目標話者が異なる言語を話す場合を対象とした声質変換技術であるクロスリンガル声質変換を利用し、英語母語話者の音声を日本人学習者の音声に変換した。変換した音声の劣化を発音評価ソフトウェアで定量化した。また実際に学習者に変換音声でシャドーイングをしてもらった。ここで使う発音評価ソフトウェアは、発話を発音・アクセント・イントネーション・タイミング・総合の5つの尺度でスコア化する。前者の音声の劣化については、英語母語話者の声を学習者の声に変換する処理によって音素単位の発音スコアが低下してしまっていたが、リズムやイントネーションのスコアは維持できていることが確認できた。

シャドーイングへの適用については、英語母語話者の音声でシャドーイングをする場合をベースラインとし、変換音声を使ったシャドーイングをする場合と比較した。実験参加者の内観報告から、「自身の音高に近いので作業負荷が比較的小さくなる」という意見や、「リズムやイントネーションが取りやすい」という意見があった。一方で、「発音が分からない」という意見や「聞き取りづらい」という意見もあり、改善が必要である。

<研究項目2: バーチャル空間での実演による学習手法>

英語の言語感覚を理解するために、バーチャル空間で身体や道具を使うことで単語の意味を再現しながら学習するシステムを実装した。例えば *fondle* (なでる) という単語を理解するために、実際に猫を撫でる動作をしたり、*conjure* (魔法をかける) という単語を理解するために、魔法陣に杖をかざすなどの動作しながら覚えるものである。

このような実演を用いた英単語学習において、バーチャル空間の背景(背景オブジェクトおよび背景描画)の有無が、英単語の記憶保持に与える影響を調べた。実験のために、動作で意味を表現できる英単語で、実験参加者の事前知識の影響が少ないと考えられる英単語を選定した。まずは SVL12000 の中から超上級(レベル10-12)の英単語を抜粋し(3000語)、WordNet のカテゴリが *verb.motion*, *verb.creation*, *verb.contact*, *verb.possession*, *verb.competition*, *verb.consumption*, *verb.body* に該当し、他動詞の英単語だけを選定した(166語)。さらに、意味の重複や wordnet カテゴリのばらつきを考慮しながら、21単語を実験で利用した。実験参加者が大学生20名で、背景描画ありの条件で学習するグループ(提案条件)と、背景描画なしの条件で学習するグループ(ベースライン条件)に分けられた。結果として、学習直後の再生テスト(英単語をみて、日本語訳を回答するテスト)において、提案手法でも記憶保持が有意に向上していた。バーチャル空間での実演を伴う英単語学習において、背景描画の有無が、学習直後の記憶保持に影響する可能性が示唆された。

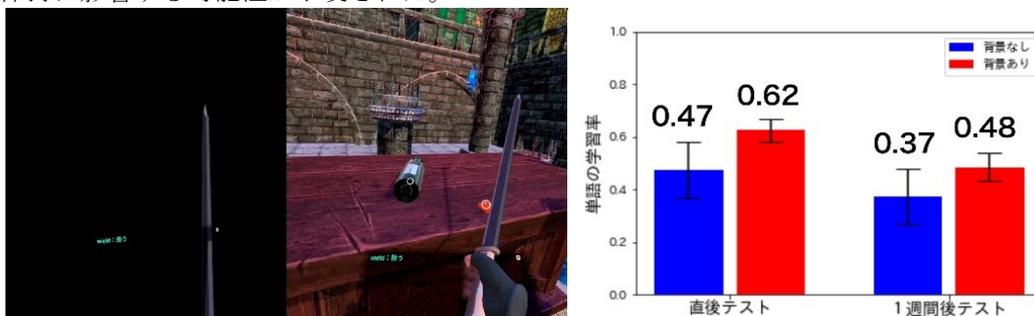


図1 左: 背景描画がない場合とある場合の違い、右: 再生テストの結果

また、実演を用いた英単語学習において、バーチャル空間でユーザーが行う動作に関する他動詞を、コーパスを使って機械的に選定することを試みた。まずは、バーチャル環境で学習可能な他動詞を調査するために、英語に精通している専門家(翻訳家あるいは英語教師)9名にアンケートを行った。アンケートは次の2つの質問である。「Q1: In general, can the transitive verb ○○ be learned using a gesture or action?」、「Q2: Can the learning of the transitive verb ○○ be helped by performing it through a gesture or action in VR space?」データセットは日本人大学生を仮定し、レベル別語彙リスト SVL12000 から上級~最上級レベルに位置するレベル7以上の他動詞1717語を使用し、参加者に重複なく分割した。その結果、Q1・Q2共に”yes”である他動詞は、874語(SVLの他動詞のうち約51%)であった。この874語を要学習他動詞とした。

次にこの他動詞を生成しうる名詞(最終的には3Dオブジェクトにリンクさせる)を探した。英語 Wikipedia のダンプファイルの一部をコーパスとして用い、Stanford Parser の構文解析機能の一つである依存関係の抽出を用いて、名詞と他動詞の関係を調査した。他動詞と名詞が一对一对応することを仮定して、第三文型(SVO)である英文のみを用いた。V-Oの関係で名詞と他動詞が繋がっている場合、その名詞と他動詞が「リンク」しているとみなした。調査対象の

名詞は、SVL に収録されている名詞のみ (7823 語) とした。実験の結果、他動詞にリンクする名詞は 6759 語 (SVL の名詞のうち約 86.4%) であった。また、コーパス中に出現した他動詞は 832 語であり、「バーチャル環境で学習可能」な他動詞のうち約 95.2% をカバーしていた。しかし、出現した名詞が 6759 語と非常に多く、また、バーチャル環境に実装可能な名詞、実装困難な名詞が混ざっており、この中からバーチャル環境に実装する名詞を選定するのが困難であった (例えば、granite は実装可能だが、landscape は規模が大きすぎる、summer は物体として出現させることができないなどの理由で実装が難しい)。

そのため、バーチャル環境に実装する名詞を選定しやすくするため、実装候補となりうる名詞を更に絞り込むための実験を行った。名詞 6759 語を対象に、Unity Asset Store⁵ の 3D/Props (小道具) カテゴリで検索を行った。検索で該当するアセットが存在しない場合、その名詞はバーチャル環境で実装不可能とみなし、実装候補から除外した。また WordNet データベースを用い、対象の名詞に下位語 (その名詞より更に具体性のある単語) が存在する場合、実装候補から除外した (例えば、carnivore は dog という下位語が存在しているため、候補から除外される)。下位語検索は、Python モジュール Natural Language Toolkit (NLTK) を使用した。実験の結果、Unity Asset Store にアセットが存在し、WordNet における下位語が存在しない名詞は 586 語存在した。この中から、他動詞のリンク数が 5 以上で、かつ存在するアセットがその名詞の意味に則しているものから 5 つ選択し、実装候補とした。選出された名詞の一部を図 2 左に示す。

バーチャル物体として実装し、バーチャル環境を作成した。honey 以外の名詞は、Unity Asset Store の無料アセットを使用し、honey は 3D モデルを自作した。各物体のフィードバックは、フィードバックにより動作が誘導されることを避けるため、「コントローラのトリガーを引くとレーザーが発射される」、「物体が重力に従い落ちる」等最低限のフィードバックのみを実装した。4 名の参加者を対象に実験を行った。動作を行う際は、その動作を口頭で宣言しながら行うこと、制限時間は設定せず、参加者が行動終了を宣言するまで実験を継続した。実験により物体から想起された行動とリンク他動詞を比較した結果を図 2 右に示す。実際にユーザの行動として現れたリンク他動詞はどの物体も半数以下だった。また、リンク他動詞の中に無い動作が、全ての物体において確認された。また、実験で想起された動作の中には、「barricade を振り回す」等、実世界では起こりくい動作が確認できた。実験によりユーザから想起されたリンク他動詞は少なかった。リンク他動詞の中には、granite に対する excavate (掘り出す) など、適切な文脈を経ないと出てきにくい他動詞があることが確認された。これらの文脈をすべて考慮してバーチャル環境に実装するのは難しい。物体同士を組み合わせる起こる動作 (honey を「使って」猫に食べさせる等) も見られた。提案手法による他動詞と名詞の関係性の網羅だけでは、学習する他動詞を選定するための判断材料として不十分であることが考えられる。最初にバーチャル環境に実装する物体を決定して、その物体から想起される動作をアンケート等で集計するなどの別のアプローチを組み合わせる必要がある。

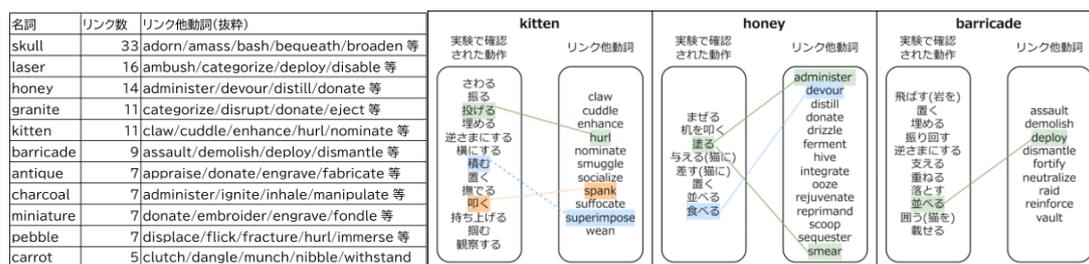


図 2 左 : 名詞とそれに紐づく他動詞の例、右 : 実験でユーザが行った動作と他動詞の対応

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 宮澤要二, 福嶋政期, ハウタサーリ アリ, 苗村 健
2. 発表標題 英単語学習時の実演効果を誘発するVRアプリケーションのデザイン ~ 背景描画の効果 ~
3. 学会等名 信学技報, vol. 121, no. 221, MVE2021-29, pp. 19-24
4. 発表年 2021年

1. 発表者名 阪本啓悟, 福嶋政期
2. 発表標題 IVRにおけるユーザーの動作に基づく学習語彙の生成手法に関する基礎検討
3. 学会等名 信学技報, vol. 122, no. 440, MVE2022-57, pp. 50-55
4. 発表年 2023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
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------