

# シャドーイングに基づく言語学習者の英語音声産出・知覚能力に 対する多角的分析

梶原 卓弥<sup>†</sup> 朱 伝博<sup>†</sup> 齋藤 大輔<sup>†</sup> 峯松 信明<sup>†</sup> 中西のりこ<sup>††</sup>

<sup>†</sup> 東京大学大学院工学系研究科

<sup>††</sup> 神戸学院大学グローバル・コミュニケーション学部

E-mail: <sup>†</sup>{kuniyara,zhuchb,dsk\_saito,mine}@gavo.t.u-tokyo.ac.jp, <sup>††</sup>nakanisi@gc.kobegakuin.ac.jp

**あらまし** シャドーイングはモデル音声を聴取し、遅れずに復唱する認知タスクの高い複合的な訓練方法である。聞き取りが不十分であれば発声が崩れ、聞き取りができていても発音を誤って習得していれば、発声はモデル音声から逸脱する。本研究では複合タスクであるシャドーイング課題の構成を工夫することで、学習者の聞き取り能力、モデル音声に対する模倣能力、更には、習得した音韻表象を（シャドー訓練後も）保持できているか否かなど、学習者の能力を多角的に分析する方法を提案する。本分析方法を、神戸学院大学グローバル・コミュニケーション学部の1,2年生を対象として導入した。本報告では、2021年度前期（夏学期）に全員を対象として導入されたシャドーイング課題（隔週の宿題）、更には、夏休みに（42日間毎日）希望者を対象として導入されたシャドーイングマラソンにおける分析結果を報告する。更に、シャドーイングの継続的な訓練が、学習者の産出や知覚に与える影響について論じる。

**キーワード** シャドーイング、英語学習、音声知覚、音声産出、ポステリオグラム、韻律的特徴、DTW

## Shadowing-based multi-faceted assessment of language learners' ability of production and perception of English speech

Takuya KUNIHARA<sup>†</sup>, Chuanbo ZHU<sup>†</sup>, Daisuke SAITO<sup>†</sup>, Nobuaki MINEMATSU<sup>†</sup>, and Noriko  
NAKANISHI<sup>††</sup>

<sup>†</sup> Graduate School of Engineering, The University of Tokyo

<sup>††</sup> Faculty of Global Communication, Kobe Gakuin University

E-mail: <sup>†</sup>{kuniyara,zhuchb,dsk\_saito,mine}@gavo.t.u-tokyo.ac.jp, <sup>††</sup>nakanisi@gc.kobegakuin.ac.jp

### 1. はじめに

外国語習熟度の自動評価として、口頭産出能力を自動評価する試みが広く行われている [1]~[3]。読み上げ音声や、状況を限定した上で学習者に発話させるなど、学習者が意図した単語列が既知となる（あるいは、精度高く音声認識できる）場合は、音響・発音分析の粒度を細かく設定し（例えば音素単位）、分析的特徴量を使って評価の自動化を検討できる。その一方で、自由発話を許した場合は音声認識精度の劣化を招き、その結果、学習者が意図した単語列を十分に正確に把握できず、分析粒度を細かくすることは得策ではない。筆者らは先行研究において [4]、準自由発話と言える学習者音声を対象として、様々な観点から手動評定された主観的スコアの自動予測を検討した。発話全体を通して得られる、全体的な音響的・韻律的・音韻的

特徴を抽出し、人間による評価値平均（全体的・主観的評価スコアの平均）を予測し、良好な結果を報告した。

本研究では、学習者が発声すべき内容を事前に示した上で発声させ、粒度の細かい特徴を用いて学習者音声を分析する。テキストとモデル音声を提示して発声・練習させれば、モデル音声へ接近することが期待されるが、その後、時間をおいてテキストのみで読み上げさせると、モデル音声から逸脱する様子を分析できる。また、音声技術を用いた外国語教育支援は、発音評価や発音誤り検出（スピーキングテストの自動化）が主であったが、本研究では、外国語音声を聴取時にどの部分で聞き取りが妨げられたのか、即ち提示音声に対する聴解（知覚）の崩れを時系列として抽出することも試みる。

具体的な発声タスクとしてはシャドーイングを採用し、学習者が発声すべき内容は音声のみで提示する。シャドーイングは、

提示音声に対する、1) 知覚、2) 理解、3) 生成（模倣）をほぼ同時に行う訓練法である。聞く、あるいは、読むだけの訓練よりも認知タスクは上がるが、通常の音声活動では、聞く、理解する、話すプロセスは同時進行する。その意味において、シャドーイングは外国語音声コミュニケーションの基礎体力作り、と考えられる。この多目的型訓練法であるシャドーイングの課題設計を工夫し、音声産出（モデル音声への接近や逸脱）の様子、音声知覚（聴解）の様子を分析する。

## 2. 先行研究

### 2.1 母語話者及び上級学習者による日本人英語音声に対する了解度の計測

多くの外国語教師は、「発音訓練の目標は母語話者のような発音ではなく、相手に伝わる発音である」と主張する。その一方で、どのような発話のどの部分はどの程度伝わっているのか、は当然、聴取者の言語背景にも依存する [5], [6]。しかし「学習者音声を聴取中に、聴取者の頭の中でどのような単語が想起されたのかを観測する」こと自体が難しい、という問題も存在する。[7], [8] では、この問題を以下の方法で解決している。学習者の読み上げ音声を、聴取者（母語話者や上級学習者）にシャドー（S）させ、その後、学習者が読み上げた原稿を見せながら再度シャドーさせる（スクリプトシャドー、SS）。なお、ここでは学習者の発音の癖を真似る必要はなく、学習者が発声した単語列を聴取者自らの発音で追追することが求められる。学習者音声の一部が聞き取れないと、シャドーは崩れるが、スクリプトシャドーは基本的に、学習者音声に同期した読み上げ音声となる。この S と SS を比較すれば「学習者音声のどこで聞き取れなくなったのか」を時系列として計測できる。具体的には、S と SS を Phonetic PosteriorGram (PPG) に変換し、両者を DTW で比較する (PPG-DTW)。[6] では、S 音声を第三者に書き起こさせ、S の書き起こしと学習者の読み上げ原稿を比較し、書き起こしに基づく（単語、音節、音素単位での）了解度を求めた。これと、（書き起こしが不要な）S と SS の PPG-DTW（単語、音節、音素を単位としたスコア計算）とを比較したところ、高い相関を得た。図 1 に母語話者 (N)、日本人の英語上級者 (J)、非日本人の英語上級者 (NN) 7 名の聴取者に対する単語単位での相関を示す。横軸が PPG-DTW による了解度、縦軸が書き起こしに基づく了解度である。個々の点は話者（学習者）を表す。聴取者に S, SS を課し、両者を自動比較すれば、様々な言語単位に対する了解度（どの程度聞き取れたのか）が計測できる。この場合聴取者が非母語話者で、S, SS に外国語訛りがあっても発話比較は同一話者内であり、訛りは精度に影響しない [6]。

図 1 に示した上級学習者（非母語話者、▲ NN1, ▲ NN2, ▲ NN3）の結果であるが、彼らのシャドー音声書き起こしより、正解聴取率は 68% であった（母語話者の場合は 85%）。つまり、聞き取りに困難を抱えながらシャドーした聴取者の了解度も、S, SS を比較することで精度高く求まっている。これは、学習者にモデル音声（母語話者音声）を S, SS させることで、学習者のモデル音声に対する了解度、即ち、聴解率（リスニング率）を計測可能であることを意味する。通常リスニングテストは、提示された音声の意味理解を問う（選択肢）問題が多い。提示

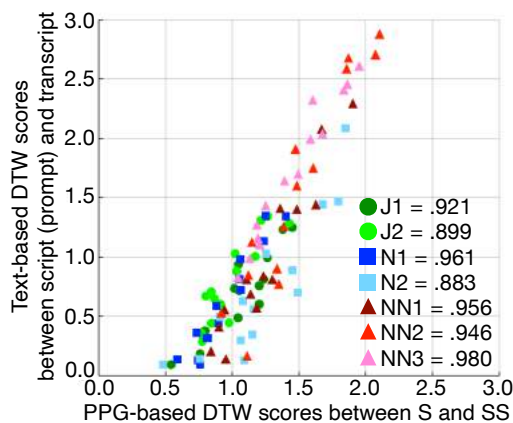


図 1 S,SS による了解度の計測 [6]

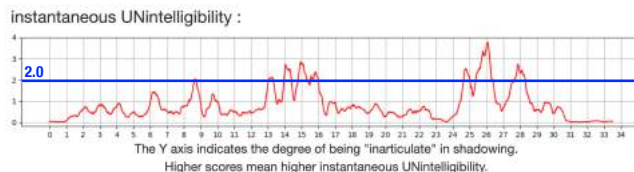


図 2 PPG-DTW(S,SS) による瞬時的了解度の時系列分析 [9]

音声のどこで聞き取りが阻害されたのかを問うている訳ではない。S, SS の比較により、モデル音声のどこで聴解が崩れたのかを時系列として、自動計測できる。

### 2.2 PPG-DTW(S,SS) に基づく瞬時的了解度の時系列分析

提示音声に対する S, SS を PPG-DTW すれば、DTW パス上の局所距離の系列は、シャドー時の発話の乱れを表す時系列データとなる [9]。図 2 はこの時系列データを、500msec を単位とした移動平均処理により平滑化した結果の例である。縦軸は DTW 局所距離であるが、およそ 2.0 を超える崩れが生じた箇所が、シャドー崩れに対応している。このように聴取中の聴解崩れは、聴取者にシャドーイングを課すことで時系列データとして容易に取得できる。類似した計測は、聴取者に書き起こさせることでも可能だが、聴取後に（書き取り中に）発話内容を推測することが不可避免的に生じることや、また、提示音声短文である必要もあり、瞬時的な了解度の計測としては不適切であると考えられる。[9] では約 30 秒ほどの音声を S, SS させている。

シャドーイングは本来、音声知覚研究において心的辞書検索過程を分析するために考案された実験タスクである [10]。それが同時通訳士や外国語学習者の訓練として利用され今日に至っているが、本研究では、シャドーイングを本来の目的である、音声知覚（聴解）プロセスの分析手法として用いている。

## 3. 発声タスクデザインと実験参加者

### 3.1 前期課題と夏休み課題の発声タスクデザイン

シャドーイングは多目的型の訓練法であり、タスク設計を工夫することで、各種目的に応じた評価のための素材を得ることができる。モデル音声 (M) に対して、シャドー (S) を繰り返すことで（即ち聴取と発声を繰り返すことで）、M の聴解がどのように向上するのかを知ることができる。よって S を複数回行い、その後、SS を課すこととした。本実験では S を三回、その後 SS を一回行わせた（図 3 参照）。なお、S 収録に関しては言い淀んだとしても再収録は行わず、Sn 毎に 1 回の収録とな

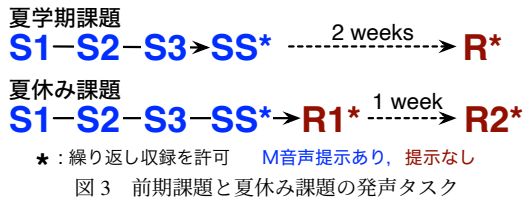


図3 前期課題と夏休み課題の発声タスク



図4 SS用Web音声収録インタフェース

る。一方、SSは $S_n$ に対する比較対象となる参照音声であるため、言い淀みなどが生じた場合は繰り返しの収録を許した。SS収録は、図4に示すように、Mの波形と再生位置、テキストが(Mの聴覚提示と共に)視覚提示される。

前期課題ではSSまでを1セッションとしたが、SSはモデル音声を聴きながらの発声であり、発話が若干不自然となる学習者がいた(言わされている感の強い学習者)。そこで夏休み課題では、SSの後に、波形は示すが音量を0としたSS(webインタフェースは図4を使うが音量が0)を、読み上げタスク(R1)として導入した。SS同様、繰り返し収録を許した。

その後、時間を置いて(前期では2週間、夏休みでは1週間)、再度音量0のSSを読み上げタスクとして行った(各々R、R2と呼ぶ)。これらのタスクは、モデル音声に対する記憶が希薄になった段階で、テキストだけを見て、SSやR1と同様の読み上げができるか否か、を検証することが目的である。

テキストを見た時に、その読み上げ音声を心的に表象する能力や表象されたイメージは、音韻表象(心的音声イメージ)と呼ばれる。SS発声時やR1発声時はMを数回聞いており、テキストを見て心的にイメージする音声はMに近づくことと予想される。一方、時間差を置いて再度テキストを見た時に、Mが心的に表象されるか否かは学習者の習熟度に依存する。その様子を、SSとRを(前期課題)、R1とR2を(夏休み課題)を比較することで検討する。なお、テキストを流暢に読めるようになるには、そのテキストを見ただけで流暢な英語音声がか心的に聞こえる能力が必須であると考えている<sup>1)</sup>。

前期課題におけるMとして、英検2級のリスニング試験で使われたモノログ音声(約30秒)を、話速を0.8倍に落として使用した。夏休み課題も同様のモノログ音声であるが、42日間(=6週間)を2週間×3に分け、0.8倍、0.9倍、1.0倍の音源を2週間ずつ続けた。実験前にシャドーイングの練習とし

て、日本語音声のシャドーイングや、one, two, threeなど平易な単語列、平易な英語文音声、そして、実験で使用しなかった英検2級のモノログ音声のシャドーイングを行わせた。

### 3.2 実験参加者と実験手順

前期課題(隔週の宿題)に関しては学部1,2年生、合計111名の音声分析対象である。夏休み課題(毎日のマラソン形式の課題)は希望者を募り、最終日まで完走したのは43名であり、彼らの音声分析対象である。

前期課題の前に概要を説明し、音声収録に関する許諾を得た後に、シャドーイング練習を課し、その後、前期課題の収録となった。S1-S2-S3-SS収録後、二週間してRを収録した。この作業を16のM音声に対して行なわせた(4つのM音声×4セッション)。音声収録は宿題として課したため、学習者間で提示音声内容について情報共有する恐れがある。そのため16種類のMは、原則学習者間で異なるものとしたが、うち4つは共通とした(セッション毎に共通M音声一つある)。本稿では、第一、第二セッションの共通M音声(以下、タスクA、Bとする)に対する分析結果を報告する。

前期課題の後、夏休み課題である毎日のシャドーイングマラソンを実施した。毎日4つのM音声×42日間(=168音声)のシャドーを繰り返すが、全参加者共通のタスクとした。継続的に学習することの効果を見るため、初日(Day01)とDay23、及び、Day20とDay42は同じ内容とした。Day23は本来話速0.9倍の期間であるが、話速0.8のDay01を、同様に、Day42は話速0.9倍のDay23を実施した。Day01 vs. 23, Day20 vs. 42の比較の他、毎日の4つのM音声に対する分析結果を通して、シャドーイングを継続する効果について論じる。

なお、本実験の参加者は事前にVersant Test[11]を受けている。総合スコアをCEFR(A1~C2)に変換すると、A1が64人、A2が45人、B1が2人であった(B2以上の学生はいなかった)。即ち、習熟度の偏りが非常に強い学習者群となっている。

## 4. 前期課題に対する分析と考察

### 4.1 学習者の音声産出に関する分析と考察

S1-S2-S3と繰り返すことでMへ接近し、SSで更に接近し、2週間後のRはMから逸脱することが予想される。 $S_n$ とMの距離 $|S_n - M|$ は、PPG-based DTWの最適パスをMから得られた音素境界で区分し、音素毎にDTWスコアを計算し、その平均値を求めた(phoneme-based DTW, pDTWと呼ぶ)。表1に各タスクでのpDTWの平均を示す。収録は、S1-S2-S3-SS-Rの順で行われたが、 $n$ 回目の収録と $n+1$ 回目の収録とで、Mへの距離が有意に減少/増加したのかをt検定した結果も示す。いずれのタスクにおいても、S1, S2, S3と距離は縮まるが、S2→S3では強い有意差はなく、シャドーを3回以上繰り返す効果は今回の収録では低い。一方、読み上げ原稿を参照させたSSではMへ一気に接近するが、Mを提示しないRは、有意に逸脱する。原稿無しでMを3度聞かせたシャドー音声(S3)よりも、2週間後にM無しで読ませたRの方がMには近い。聞き取りに苦勞すると共に、(仮に何が発声されたのか単語が同定できなかったとしても)耳からの情報を自らの調音運動に反映させることに苦勞していることが窺える。言い換えれば、目からの

(注1): シャドーイングをメロディーのハミング練習に例えると理解し易いかも。新曲のメロディーとその楽譜を同時に示してハミング練習を数回行った二週間後に、楽譜だけでハミングできるのか(楽譜からメロディーが心的に聞こえるのか)という問題である。楽譜を英語テキストとしたものが本実験である。

表1 モデル音声 (M) への接近と逸脱

数値が低いほど接近度合いが高い。

タスク	S1-M	S2-M	S3-M	SS-M	R-M
A	1.91	1.85	1.83	1.63	1.67
	—	p<0.001	p<0.1	p<0.001	p<0.005
B	1.83	1.77	1.75	1.55	1.61
	—	p<0.001	p<0.05	p<0.001	p<0.001

表2 スクリプトシャドー音声 (SS) への接近

数値が低いほど接近度合いが高い。

タスク	S1-SS	S2-SS	S3-SS	R-SS
A	1.68	1.54	1.50	1.20
	—	p<0.001	p<0.001	p<0.001
B	1.43	1.32	1.27	1.17
	—	p<0.001	p<0.001	p<0.001

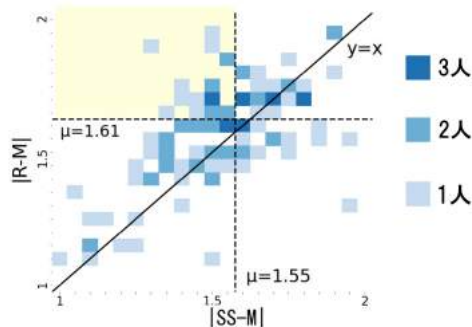


図5 タスク B の |SS-M| と |R-M| の分布 (破線はそれぞれの平均)

文字情報に頼って発声していると言える。

|SS-M| から |R-M| への増分、即ち、M を未提示とする影響についてより詳細に検討する。テキストを見て心的に表象される内的音声音が音韻表象であるが、事前の M 提示などがなければ、そのテキストに対する自らの発音 (L2 発音) が音韻表象となる。Sn, SS と 4 回 M をシャドーする (真似る) ことで、音韻リハーサルを繰り返し、学習者の音韻表象が M に近づく効果が期待される [12], [13]。しかし、二週間後の R の結果 (表 1) は、学習者の音韻表象が M から逸脱する様子が見られる。

図 5 に横軸 |SS-M|、縦軸 |R-M| とし、タスク B の学習者数ヒストグラムをカラーマップの形で示している。全体の傾向としては表 1 が示す通り |R-M| > |SS-M| であるが、|R-M| < |SS-M| となる (SS より R の方が M に近づく) 学習者も少なくない。これらを聴取すると、(SS とは言え) モデル音声を聴取しながらの発声であり、モデルに追従することに注意が割かれたような発声 (言わされている感の強い発声) が見受けられた。夏休み課題として、SS の直後に R1 を追加した理由がこれである。

さて、図 5 の左上に位置する学生は興味深い。|SS-M| の分布からは上位に位置する (M より近い) が、|R-M| の分布からは下位に位置する (M からより遠い) 学生である。M を数回聞くことで、M と類似した音韻表象を獲得し、自身も M に近い発音をしているものの、時間を置いてテキストのみで読ませると、大きく逸脱している (音韻表象が元に戻る) 学習者である。言い換えれば、モデル音声を真似することは得意だが、その音声イメージを維持することが難しい学習者と解釈できる。

## 4.2 学習者の音声知覚に関する分析と考察

S1, S2, S3, R の各音声を SS を基準に pDTW を計測した結果を表 2 に示す。SS は、内容を把握した上でのシャドーであり、表 2 は、シャドーイングの 3 タスクのうち、1) 知覚、2) 理解がどの程度出来ているのかを定量化していると解釈できる (第 2.1 節参照)。M を基準として計量をした表 1 と比べて有意に小さいが、表 1 と異なり、S2→S3 の場合でも距離減少が有意であることは興味深い。今回の収録では、M への接近に関しては 3 回以上シャドーすることの効果は少ないが、M の聞き取りに関しては、3 回以上聞くことの効果が見える。

シャドーイングは、外国語教育に導入された当初、リスニング力向上のトレーニング教材として導入されている [14], [15]。その後、発音の訓練 (M に近い音韻表象の獲得) としての効果が検討されるに至っている [12], [13]。図 3 に示したタスクデザインでは、SS 時にテキスト提示しており、これを正解として、聞き取れなかった/聞き間違えていた箇所を学習者は確認できる。一方、発音に関しても基準となる M を数回提示している。しかし表 1 に示した |S3-M| と |R-M| のスコア差異は、学習者が一人で、M と自らの発音のズレを検出、矯正できないことを示唆している。第三者の適切な発音指導を与えない状況でシャドーを繰り返すことが、どれほど学習者の発音を改善させるのだろうか? どれほど学習者の音韻表象を変えるのだろうか? この疑問に答えるため、42 日間に渡る (約 1 時間/日) のシャドーイングマラソンを夏休み課題として企画し、希望者を募った。かなりのタスク量であるが、全学生に実施した前期課題は「自身のシャドー能力向上が把握できる教材」として人気の学習教材となっており、夏休み課題も多く多くの学生が参加してくれた。

## 5. 夏休み課題に対する分析と考察

### 5.1 新たな評価項目の導入

前期課題の分析で用いた、2 発声の PPG-DTW より計算される pDTW は、当該発話間で発音 (舌や唇の動き) の違い、すなわち調音制御の違い、分節的特徴の違いがスコア化されている。外国語発音を指導する場合、個々の音の生成のみならず、音を超えた韻律 (超分節的特徴) の指導も必要である。そこで本節では、韻律に関しても 2 発話間比較を行った。

声のパワー、ピッチ、継続長に関する制御の類似性を二発話間で比較した。まず、分析対象の音声のパワーパターンとして、各フレームの RMS (Root Mean Square) を時系列として求めた。ピッチパターンも同様、各フレームの基本周波数を時系列として求めた。その後、二発話に対して強制アライメント処理を施し、音素境界を得ると同時に、二発話間で PPG-DTW を行い、発話間の時間対応付けを求めた。二発話のパワー制御における類似性は、PPG-DTW 結果から対応するフレーム対を検出し、ともに母音フレームである対のみを取り上げて、相関係数としてスコア化した。ピッチ制御における類似性も同様にして、時間合わせをした二発話間でピッチパターンの相関係数でスコア化した。継続長については各フレーム毎に定義できるものではないため、上記とは異なるスコア化を採択した。強制アライメント処理の結果、各音声に対して、第 1 強勢、第 2 強勢、弱勢を区別した 45 種類の母音に対して境界位置が求まる。その結

果より、45種類の母音の平均継続長を求め、二発話間でこれらの相関係数を計算してスコア化した。相関計算時の分布図は、パワー、ピッチに関しては(母音の)フレーム対の総数がサンプル数となるが、継続長については母音種類数がサンプル数となる。以下、調音制御、パワー制御、ピッチ制御、継続長制御のことを、簡単のため、調音、大きさ、高さ、長さ、と呼ぶ。

## 5.2 調音、大きさ、高さ、長さの分析結果

第4.1節で示したタスクA, Bは、いずれも一種類のM音声に対するシャドー音声の結果であるが、夏休み課題では毎日4つのM音声に対してシャドーイングを行っている。例えばある学習者のDaynの調音スコアは、4つのM音声に対する調音スコアの平均値である。大きさ、高さ、長さも同様、4音声の平均値を当該学習者のスコアとしている。以下、同じM音声を使用したDay01 vs. 23, Day20 vs. 42を、調音、大きさ、高さ、長さ制御において比較する。表3, 表4に調音(pDTW)、大きさ、高さ、長さ(相関)の学習者平均を示す。日によって参加者数が変わるので、Day01 vs. 23, Day20 vs. 42は、32名、36名が分析対象となっている。スケジュールの都合上、Day36以降はR2が収録できていないため、Day42のR2は除外している。

表1にてMへの接近と逸脱の様子を議論したが、同様の傾向は表3にも観察される。Sを繰り返すことでMに近づくが、テキストを提示するSSで一気に近づく。Mを3回聴取した|S3-M|よりも、一週間後のテキストのみを参照した|R2-M|の方が低くなっている。表2にてSを繰り返すことで聴解精度が上がる様子が示されたが、同様の傾向は表4にも観察される。また、SS直後のR1とSSとの差異は極めて小さくなるが、一週間後のR2はSSから逸脱する様子が示されている。

## 5.3 学習者の音声産出に関する考察

表3のDay01を23と、Day20を42と比較し、シャドー訓練を継続する効果について考察する。Mへの接近については、|Sn-M|は01→23, 20→42において減少しているが、肝心の|R1-M|, |R2-M|においては、シャドー訓練を継続しても、Mに近づく効果は有意には観測されなかった。適切な教示がなければ、三週間ほど毎日1時間シャドー訓練を継続しても、調音制御を変える効果を期待することは難しいことが分かる。

Mへの接近の様子について、韻律の面から考察する。大きさ、高さ、長さ、いずれにおいても、Sn (n=1, 2, 3)とSを繰り返す明確な効果は、訓練を継続する前(01, 20)も後(23, 42)も極めて僅かである。S3からSSとテキストが表示されると、一部の相関は上がるが、調音制御(pDTWの大きな下落)に比べると、韻律相関の上昇は極めて限定的である。

更に、SSの直後にMなしで読み上げさせた場合(R1)、調音と韻律では異なる様相を呈した。調音の場合SSからR1になっても、Mとの距離は変化せず、また、シャドーを継続して行う前後でも変化は乏しい。韻律の場合も同様にSSの直後にR1を課しているが、R1となることで相関が下落する。特にピッチ制御における下落が激しい。類似した現象は日本語学習者のシャドーイングでも観察されている。日本語教育の場合、日本語の単語アクセントが高さアクセントであることを明示的に教示しないことは少なくない。このような場合、シャドーイングを課せば適切な箇所アクセント(ピッチの下落)を生成するが、その後、読み上げさせると不適切なアクセントとなる例が報告

表3 モデル音声(M)への接近と逸脱

調音は数値が低いほど、韻律は数値が高いほど接近度合いが高い。

		S1-M	S2-M	S3-M	SS-M	R1-M	R2-M
調音	Day01	1.83	1.76	1.74	1.52	1.55	1.59
	Day23	1.63	1.61	1.61	1.52	1.54	1.53
	Day20	1.82	1.78	1.75	1.51	1.51	1.52
	Day42	1.69	1.68	1.65	1.50	1.50	—
大きさ	Day01	0.60	0.61	0.60	0.62	0.61	0.60
	Day23	0.62	0.62	0.63	0.63	0.61	0.62
	Day20	0.56	0.56	0.59	0.59	0.55	0.55
	Day42	0.58	0.58	0.61	0.61	0.56	—
高さ	Day01	0.51	0.51	0.52	0.51	0.32	0.31
	Day23	0.63	0.60	0.60	0.59	0.42	0.36
	Day20	0.61	0.62	0.61	0.62	0.44	0.39
	Day42	0.65	0.65	0.66	0.64	0.47	—
長さ	Day01	0.56	0.59	0.58	0.65	0.60	0.58
	Day23	0.61	0.65	0.64	0.65	0.61	0.59
	Day20	0.62	0.62	0.64	0.67	0.60	0.58
	Day42	0.64	0.65	0.68	0.70	0.64	—

表4 スクリプトシャドー音声(SS)への接近

調音は数値が低いほど、韻律は数値が高いほど接近度合いが高い。

		S1-SS	S2-SS	S3-SS	R1-SS	R2-SS
調音	Day01	1.40	1.28	1.24	0.97	1.16
	Day23	1.11	1.02	0.99	0.90	1.08
	Day20	1.35	1.27	1.24	0.93	1.06
	Day42	1.16	1.10	1.05	0.90	—
大きさ	Day01	0.66	0.69	0.70	0.73	0.68
	Day23	0.72	0.73	0.74	0.75	0.70
	Day20	0.67	0.70	0.69	0.73	0.68
	Day42	0.70	0.72	0.72	0.73	—
高さ	Day01	0.54	0.59	0.61	0.53	0.48
	Day23	0.70	0.72	0.72	0.61	0.50
	Day20	0.66	0.68	0.69	0.63	0.55
	Day42	0.70	0.72	0.74	0.63	—
長さ	Day01	0.61	0.62	0.64	0.74	0.70
	Day23	0.71	0.74	0.73	0.79	0.73
	Day20	0.65	0.68	0.69	0.73	0.71
	Day42	0.70	0.74	0.75	0.78	—

されている[16]。韻律の場合、テキストが表示されても、そこに明示的に答えが示されている訳ではなく、また、英語の韻律制御についての知識が不十分であれば、モデル音声の韻律制御(韻律構造)の何を模倣すべきなのかが分からないため、Mの消失と共に、学習者による韻律制御は不安定なものとなる。Mの消失と共に調音制御が乱れる様子を図5の左上に示したが、韻律制御の場合、多くの学生がここに位置すると推察される。シャドーイングを繰り返すこと、それを継続的に実施することで、テキスト読み上げにおける調音制御、韻律制御がどう変わるのか(学習者自身がテキストに対して持つ音韻表象がどう変わるのか)を検討したが、適切な音声指導が与えられずに、ただ「聞いたその場で真似る」ことを繰り返すだけでは、学習者の音声産出を改善させることは難しいと言える。

なお[17]~[19]では、適切なモデル音声(Golden speaker)を選ぶことで発音学習の効率が上がることを報告している。発音学習は「提示音声刺激を調音運動へ転化する作業」と解釈できるが、この作業精度を高めるモデル音声(真似しやすい声)を提供することで、発音学習が改善される可能性はある。

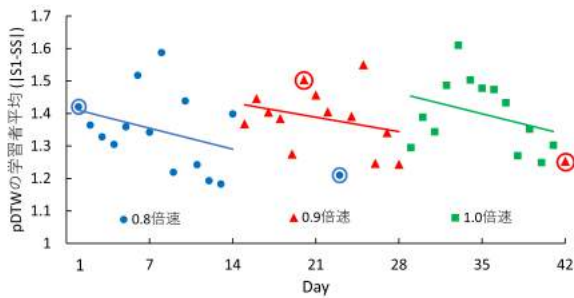


図6 継続的なシャドーイングによる聴解崩れ (|S1-SS|) の減衰

#### 5.4 学習者の音声知覚に関する考察

音声産出には効果が見られなかったが、音声知覚（聴解）の方は、シャドーイングを継続的に行うことで大きく改善する結果となった。表4のDay01→23, Day20→42で、pDTWスコアは大きく減少している。Day23, Day42の|S1-SS|は、Day01, Day20の|S3-SS|よりも小さい。S1から円滑にシャドーできている様子を窺うことができる。これは非常に大きな改善である。当初議論されていたように、シャドーイングの主たる効果は聴解力を上げることにあると言える。シャドーイングを継続することで、|Sn-M|は確かに減少している（表3）が、これは、SnがMに近づくのではなく、SnはSSに近づき、その結果Mにも近づいていると解釈する方が適切であろう。そしてSSやR1, R2はシャドーイングを継続しただけでは、Mに近づくことが困難であることは第5.3節で述べた通りである。

表4では、|Sn-SS|変化の様子を韻律の面からも分析している。大きさ、高さ、長さのいずれにおいても、また、Day01, 23, 20, 42のいずれにおいても、|S1-SS|よりも|S3-SS|の方が相関が高い。すなわちSを繰り返すことで、韻律的にSSに近づいている。同様のことは表3の|S1-M|と|S3-M|でも言えるが、表3より表4の方が相関は高い。S1, S2, S3とSSに近づいた直後の読み上げ（R1）では、大きさ、長さでは、常に|S3-SS|<|R1-SS|となるが（すなわち、Mよりもテキストを参照できる効果が高い）、高さに関しては逆の結果となった。Mとの比較を行った表3でも、M音声の消失が高さ制御に大きな影響を与えたが、SSとの比較でも同様の結果となった。高さ制御に関して学習者が身につけた音韻表象（心的イメージ）は、自己流の音韻表象となっている、あるいは、不安定な音韻表象となっていると推察される。一週間後のR2では、いずれも相関はR1時の値より小さくなった。

#### 5.5 42日間を通しての聴解力の向上

以上の考察より、シャドーイングによる学習効果は主として聴解力の向上にあると言えるが、42日間を通してどのように向上したのか、|S1-SS|（但し、各日の第四課題）の学習者平均値の推移と線形回帰した結果を図6に示す。M音声の意味的難易度にはばらつきがあり、Day01 vs. 23, Day20 vs. 42のような厳密な比較はできないが、2週間ごとに話速が0.8倍、0.9倍、1.0倍と変化することも踏まえて考えると、シャドーイングを継続することで、S1の崩れが減っていく様子が分かる。

(注2)：本研究ではシャドーイングをM音声を提示するだけで行っているが、まずテキストを示して読解させ、その後で、M音声に合わせてシャドーさせる方法も広く行われている。この場合内容が既知となり、聴解力向上は難しいだろう。

## 6. まとめと今後の課題

多目的型訓練法であるシャドーイングにおけるタスクデザインを工夫し、学習者の音声産出と音声知覚の様子を、調音制御と韻律制御に分けて分析する手法を提案し、大学1,2年生を対象に収集したデータを用いて検討した。その結果、シャドーイングの主要な効果は聴解力の向上にあることが示された。また、調音制御と韻律制御の違いが明確になった。現在、様々な英語発音に対する聴解力を上げるべく、世界諸英語音声を対象としたシャドーイングを実施している。シャドー崩れを通して各学習者の聴解脆弱性をモデル化すれば、そのモデルを用いて各学習者毎に、より適したシャドーイング素材の自動選択が検討できるだろう。シャドーイングを通して音声産出の改善を狙う場合は、真似しやすい話者（Golden Speaker）の採用や、図4に、可視化された韻律パターンを追加するなどを検討したい。

## 文 献

- [1] L. Chen, J. Tao, S. Ghaffarzadegan, Y. Qian, "End-to-end neural network based automated speech scoring," *Proc. ICASSP*, pp.6234-6238, 2018
- [2] Y. Xiao, F. Soong, W. Hu, "Paired phone-posteriors approach to ESL pronunciation quality assessment," *Proc. INTERSPEECH*, pp.1631-1635, 2018
- [3] Y. Qian, P. Lange, K. Evanini, R. Pugh, R. Ubale, M. Mulholland, X. Wang, "Neural approaches to automated speech scoring of monologue and dialogue responses," *Proc. ICASSP*, pp.8112-8116, 2019
- [4] 桐原, 藩, 齋藤, 峯松, 齊藤, ボイクマン, "英語学習者と日本語学習者音声に対する多角的評価に関する実験的検討", 音講論, 2-2Q-9, 2021
- [5] C. Zhu, Z. Lin, N. Minematsu, N. Nakanishi, "Analyses on instantaneous perception of Japanese English by listeners with various language profiles," *Proc. The Phonetic Society of Japan General Meeting*, pp.26-31, 2020
- [6] C. Zhu, R. Hakoda, S. Saito, N. Minematsu, N. Nakanishi, T. Nishimura, "Multi-granularity annotation of instantaneous intelligibility of learners' utterances based on shadowing techniques," *Proc. ASRU*, 2021 (to appear)
- [7] Y. Inoue, S. Kabashima, D. Saito, N. Minematsu, K. Kanamura, Y. Yamauchi, "A study of objective measurement of comprehensibility through native speakers shadowing of learners' utterances," *Proc. INTERSPEECH*, pp.1651-1655, 2018
- [8] Z. Lin, R. Takashima, D. Saito, N. Minematsu, N. Nakanishi, "Shadowability annotation with fine granularity on L2 utterances and its improvement with native listeners' script-shadowing," *Proc. INTERSPEECH*, pp.3865-3869, 2020
- [9] N. Minematsu, R. Hakoda, C. Zhu, N. Nakanishi, T. Nishimura, D. Saito, "Objective measurement of instantaneous intelligibility of L2 utterances based on shadowing," *Proc. ISAPh*, 2021
- [10] W. D. Marslen-Wilson, "Speech shadowing and speech comprehension," *Speech Communication*, 4, pp.55-73, 1985
- [11] Versant English Speaking Test <https://www.versant.jp>
- [12] 門田修平, シャドーイング・音読と英語習得の科学, コスモビア, 2012
- [13] 門田修平, シャドーイング・音読と英語コミュニケーションの科学, コスモビア, 2015
- [14] 玉井健, "シャドーイングの効果と聴解プロセスにおける位置付け", 時事英語研究, 36号, pp.105-115, 1997
- [15] 玉井健, "リスニング向上におけるシャドーイングの効果について", 日本通訳学会年次大会, 基調講演, 2002
- [16] 阿栄娜, 林良子, "日本語学習者によるアクセントの意識化を伴う発音訓練の効果", 「日本語音声コミュニケーション研究のこれまでとこれから」論文予稿集, 2015
- [17] K. Probst, Y. Ke, and M. Eskenazi, "Enhancing foreign language tutors - In search of the golden speaker," *Speech Communication*, 37, pp.161-173, 2019
- [18] R. Wang and J. Lu, "Investigation of golden speakers for second language learners from imitation preference perspective by voice modification," *Speech Communication*, 53, pp.175-184, 2011
- [19] S. Ding et al., "Golden speaker builder - An interactive tool for pronunciation training," *Speech Communication*, 115, pp.51-66, 2019