

# スピーキング 3 秒採点と赤ペン添削システム開発とオンライン授業 人手による主観評価から機械（AI）による客観評価へ

田淵龍二  
ミント音声教育研究所  
tabuchiryuji@nifty.ne.jp

Ryan Spring  
東北大学  
spring.ryan.edward.c4@tohoku.ac.jp

## 1 はじめに

言語能力 4 技能「読む、聞く、書く、話す」において「話す」能力測定には難点が多く、入学試験への導入が進まない。とりわけ採点者を複数用意することが困難であり、時間と経費が膨大になってしまう。根本的な問題はスピーキング学習法・教授法の遅れにある。他教科の勉強と比べて見よう。

たとえば受験生は過去問を解いたとき必ず答え合わせをし、不正解なら解説を見て間違いを正すことで自己研鑽する。しかしスピーキングでは答え合わせができない。たとえ模範解答の音声があったとしても自分の発音の修正点は自明ではない。横に英語母語話者がいればと思うだろう。つまり現状のスピーキング学習法には自習環境が欠けている。

もう一つの問題は指導者と生徒の間で、生徒の発音のどこがどのように間違っているのかを共有することが難しいことだ。たとえ波形や発音記号や口形を見せられても腑に落ちるとは限らない。また厳密な採点に複数の専門家が必要なことも、課題解決の困難さを増大させている。

現在のスピーキング学習と指導には持続可能な発展（SD: Sustainable Development）に欠けている。こうした難点を解決するために、自動採点システムの研究開発も進んでいて、あと一息である[1][2][3][4]。

そこで筆者らは AI 音声認識技術（ASR: Automatic Speaking Recognition）に即時採点を可能にする自動赤ペン添削（ASCwRP: Automatic Speaking Correction with Red Pen）を加えたサービス（ASCA: Automatic Speaking Correction Application）の開発と授業実践に取り組んだ。ASCA の設計思想を具体化したオープンサイト NatTos (<https://www.mintap.com/nattos/>) の公開を受けて、音声指導授業を実施した。結果は思った以上に良好で、従来の難点とされた透明性、客観性、機会均等性、即時性で効果が確認された。また学習性、説明性、納得性、共有性、指導性を解決

する方向性が見えてきた。特にオンライン授業の受講者の多くが積極的に参加し、満点を取る（通じる）まで何度も自ら工夫しながら、授業時間終了後にも挑戦する様子が見受けられた。

以下に詳しく述べる。

## 2 ASCA を具体化した NatTos の概要

スピーキング自動赤ペン添削（ASCwRP）の仕組みはシンプルで、AI 音声認識技術（ASR）が出力したテキスト（書き起こし文）を課題テキストと比較し、双方で一致しないところを抽出する流れである。赤ペン添削の様子を図 1 に示す。



図 1 赤ペン添削の様子

上段の英文が読み上げ用の文（課題文）で、下段が受験者の音声を ASR で書き起こしたテキスト（解答文）である。赤字の文字が一致しない部分を示している。この例からは発音上の 3 つの弱点が指摘できる。

1. 言い出しの At が弱すぎたのだろう
2. 音素 TH が S と聞き取られている
3. 音素 L が R と聞き取られている

結果として“沈思”が“魅惑的？に歌う”になったことがわかる。採点と添削は直感性が高いので無用かもしれないが念のため、図 1 を参考に採点と添削と学習指導について説明する。

採点は「正しさ」と「なめらかさ」で評価している。「正しさ」は赤文字の比率を基礎にした 10 段階

評価で、満点が 10、最低が 1 となる。図 1 の例では前半ができたが後半が全滅したので 3 点となつた。

「なめらかさ」は発声単語数を発声時間（秒）で割ったもので、単位は wps (words per second)。最大は 3.1、有効桁数は小数点第 1 位で抑えてある。

添削結果は課題文と解答文に赤字で示してある。課題文の赤字は解答文にない文字（文字列）で、解答文の赤字は課題文にないことを示している。双方を照らし合わせることで「言い間違い」や「欠落」や「余分」を知ることができる。

総評は右下にあるアイコンである。ここでは通じなかつたので、「ウツ」との顔文字になっている。満点（正しく通じた）では○が表示される。

指導は受験者の個人差が出やすいが、アルファベットと音声（音素）の区別がつく初級以上であれば理解できるだろう。このあたりは学力（学年）に合わせた指導法の研究が待たれる。今回の授業対象は大学初級で、説明抜きで理解されたようだ。

### 3 NatTos による発声練習の仕組み

ここでは課題文の提示から、発声・添削までの流れを述べる。

開始ボタンをタップすると課題文が提示される。と同時に模範音声が流れる。その様子を図 2 に示す。

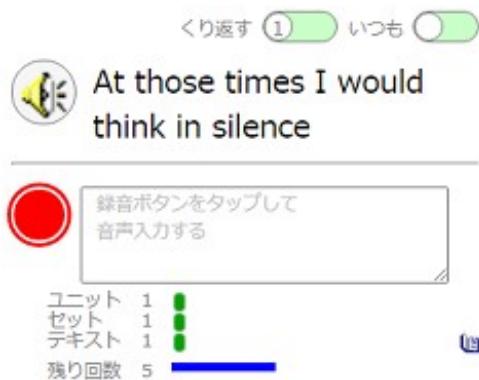


図 2 課題文提示の様子

模範音声を聞いたあとで録音ボタン（赤丸）をタップする。録音が始まるので課題文を読み上げる。読み上げを終えるか無音期間が長く続くと録音状態が自動的に解除され、添削が開始される。3 秒しないうちに先に示した図 1 のような結果が示される。

入学試験や資格試験であれば次の課題に進むことになる。しかしウェブアプリ NatTos は発声練習用として開発されている。練習用と受験用で大きく異なる点は反復性であろう。受験の解答は 1 回限りだが、

練習では何度も正解するまで挑戦することができる。工夫しながら発声・採点・添削を繰り返すことが成長につながるからである。NatTos では最大 5 回まで挑戦できる。さて、赤ペンのついた部分を注意すれば良い事がわかつても、どのように発声すれば良いのかわからない。そこで模範音声を再生できるようになっている。それがスピーカマークである。また何度か続けて聞いて確かめたいこともある。それが図 2 の右上にある「繰り返す」ボタンだ。最大 5 回まで連続再生できる。模範音声を聞きながら重ねて音を合わせるように（シャドーイング）するのが上達のコツである。また人の一息の発声と聴覚作動記憶（音韻ループ）の時間幅はともに 2~3 秒[5][6]なので、模範音声をフレーズに分けてシャドーイングできるようにしてある。それが図 1 上段の背景色（薄い青と緑）で区切られたブロックである。

### 4 赤ペン添削による学習履歴

発声と添削の様子は克明に記録され、先生も生徒もリアルタイムで授業中に閲覧できる。採点の生徒への即時フィードバックは学習を促す。他方、視覚化された学習履歴の統計処理結果は教員の形成的評価（Formative Assessment）として即時指導に利用できる。そのフィードバック機能をいくつか紹介する。

#### 4.1 クラス履歴

データマイニングによるスピーキングテストの履歴俯瞰では、正しく発音できなかった課題文や表現が一望できる（図 3）。



図 3 課題文提示の様子（抜粋）

右端のコインと小判のアイコンがそれぞれ正しさとなめらかさの度合いを示す。左端の赤と緑のバーが出来具合を示す。赤が多いほど不出来だったことを示す。濃淡のある青い背景色を持つ文字列は課題文で、色の濃いところほど正しい発音にならなかつたことを示している。付録 A にテスト課題全文と履

歴全体図を掲載した。

テキストマイニングによる語彙クラウドでは、生徒が頻繁に失敗した語彙がひと目でわかる（図 4）。

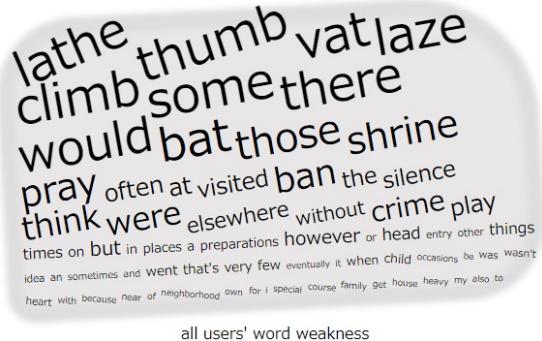


図 4 語彙クラウド例

発声に何度も失敗した語彙が太字で大きく表示される。たとえば中断左端の pray を選ぶと、図 5 のように表示が変わる。pray は 344 回（背景赤と緑）の発声があり、通じたのは 28%（緑字）だとわかる。

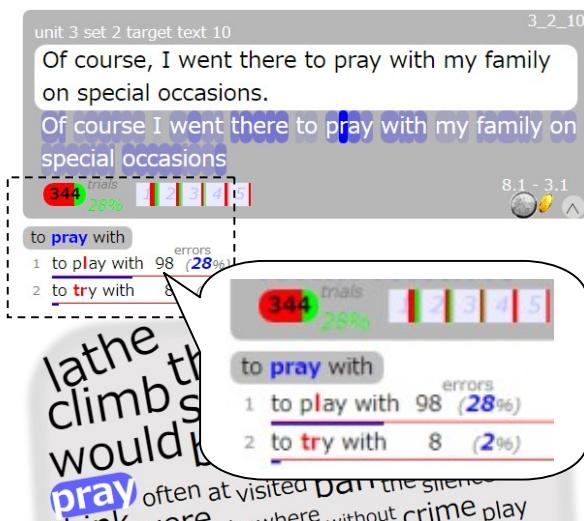


図 5 失敗パターンチャート例

さらにその下には対象語 pray の発音間違い（話者の意図と異なる ASR の聞き取り）が多かった失敗例が 2 つ示されている（破線内拡大図）。音素 R を L と発声した例が 28%にも及んでいたことがわかる。また語頭子音の曖昧さも一定量見て取れる。

## 4.2 生徒履歴

受験者の個別履歴を図 6 に示す。受験者 1 は 1 回目のテストで失敗し、2 回目のテストの 3 回目の発声で成功している。受験者 2 は 1 回目のテストの 2 回目の挑戦で成功し、2 回目テストでは一発で成功

させている。学力差だと片付けられない情報がこの履歴には隠されている。それが赤い小さな点である。

小さな赤い点の数を比べると、受験者 1 には 2 個、受験者 2 には 5 個見つかる。赤い点の数は補助音声再生の回数を表している。行頭の赤い点は自動再生で、それ以外は受験者の手動再生である。受験者 2 は 1 回目のテストで最初に失敗したあと補助音声を 3 度再生し、その後の発声で成功させていたことがわかる。母語話者の音声を参照しそれを真似て発話しようとの意志がうかがえる。それに対し受験者 1 は自らの記憶にある音声と発声知識だけを参照して失敗を繰り返したと推察される。

受験者 1



受験者 2

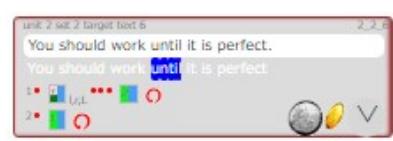


図 6 受験者ごとの発声履歴

凡例：■ 1 回分の発声 ■ 補助音声再生 ■ の 1 列分が 1 回分のテスト ■ 成功 ■ 失敗 アイコンは左から右、上から下への時系列。

## 5 オンラインテストと指導の様子

約 100 人（3 クラス、CEFR A2~B2：初級～準上級）のオンライン授業で NatTos を発音練習ツールとして利用した。発音の講義（週 1 レッスンを 5 週）の各テーマは子音、母音、ストレス、連結・脱落、リズムであった。毎回、授業の前後に教員が自分で NatTos 上に作成した発音テストを実施した。毎回の発音テストはレッスンテーマに沿った物であり、学習履歴を形成的評価として利用した。レッスン冒頭の発音テスト（前テスト）の結果を確認し、多くの学生が困ったポイントに指導を集中した。例えば、CEFR B2 のクラスは L の音が最初からできたが R の音ができなかったため、そのクラスでは R に集中的に指導して、練習させた。一方、CEFR A2 のクラスは L と R の音を両方できなかつたため、両方を平等に指導し、その区別に焦点をおいて指導した。

また、発音の指導を実施した後、NatTos で作成した別の練習課題を学生に与え、学習履歴を確認し、

最も多い間違いを参照しながらクラスごとに助言をした。例えば、CEFR B1 のクラスでは R を発音する際に W として認識された件が多かったため、多くの学生が R を発音する場合、唇を前に出していったということが特定できたため、唇を後ろに引っ張るように指導ができ、発音の改善を図れた。クラスによって発音によるミスが異なることから、学習履歴でクラス固有の弱点を即時に確認することで、教育の改善が適切にできたと考えられる。その例として、3 クラス全体の前後テストの分析結果を表 1 に示す。第 1 回（子音）と第 2 回（母音）の発音テストの正確さを対応あり t 検定で比較した場合、上達していることが明らかになった。

表 1 前後テストでの正確さの比較 N=78

レッスン	前テスト	後テスト	t 検定
子音	5.48 (1.1)	5.76 (1.0)	-2.53, p=.015
母音	5.70 (1.3)	5.97 (1.1)	-2.29, p=.026

前テストは授業冒頭、後テストは授業の締めくくりにそれぞれ 20 分ほど実施。値は平均値 (10 点満点)。カッコ内は標準偏差。前後テストのどちらか一方だけのデータは除いた。

教員として、発音の形成的評価が即時に数多くの学生にできることが非常に好ましい状態であり、発音の指導をより良好にできる。従来の発音指導法では、全ての学生の発音を聞いて、個人アドバイスするのが大変困難である。対面式の場合でも、10 人以下の少人数クラスでもない限り、教員は全員の発音を認識するのは困難であり、コロナ禍で急にオンライン化された授業では尚更できない。また、対面式で一人一人に発音指導をすると、他の学生の練習が止まり、先生の助言が来るまで待っている時間が多い。そのため、NatTos を授業に取り入れることにより、学生の練習時間が多くなり、学生が ASR によるフィードバックを即時に確認ができる、また ASCwRP 履歴マイニングに基づいたフィードバックが教員から来るため、学生には練習時間とフィードバックの量が増えることになった。

今後の研究課題として、どのような練習文を使えば、最も発音トレーニングのためになるかを調査し、どのような発音レッスンを実施すれば、最も発音が上達するかを調査したいと考える。

## 6 おわりに

英語スピーキングテストを全自動 3 秒で採点と添

削をして返すアプリは他に例がないようであったが、ASR と ASCwRP を組み込んだ ASCA で設計したアプリ NatTos が 3 クラス 100 人規模の授業で結果を出したことで、成果と課題が浮き彫りになった。

成果は、発声・採点・添削・指導のサイクルを 1 コマの授業内で可能にしたこと、通常授業でスピーキング課題をもっと頻繁に行えるようになるだけでなく、学期試験や入学試験への応用可能性も示唆された。

特に、スピーキング学習にあっては「他者とのコミュニケーションを妨げる発音誤りを認識する必要」[4]があることから、ASCA はその課題をクリアしたと判断できた。

次に ASCA の課題を列举する。

- (1) 評価の妥当性については、人手による主観的評価との比較研究を準備中である。
- (2) 学習効果については、今回のパイロット研究をベースに来年度に精度高く検証予定である。
- (3) Chrome と Safari による ASR で同等の出力が期待できるかの検証が必要である。
- (4) 読み上げ問題には対応している一方で、自由発話問題や文生成問題[1]への対応にはキーワード生成や文法チェックなどの技術導入が必要となるだろう。
- (5) 発音評価指標とされる intelligibility (音韻的正確性; 了解性) と comprehensibility (理解可能性; 可解性) [4]に従えば、ASR は後者の comprehensibility に重きをおいていると考えられることから、ASR の選定 (どの ASR を使うか、あるいは複数の ASR が必要か) などの検討も視野に入れる必要がある。
- (6) ASCA が出力する各種履歴データに支援された教授法の考案と、授業中の指導に結びつく出力データの最適化が望まれる。
- (7) 課題文の作成にあっては、単語と短文と長文で語彙や文構造による学習効果度合や難易度 (ASR の特性) を調べる必要がある。
- (8) 特に単語については ASR の成否が採点結果に大きく影響する[1]ことから親密度や出現頻度や同音異語などへの対処が求められる。

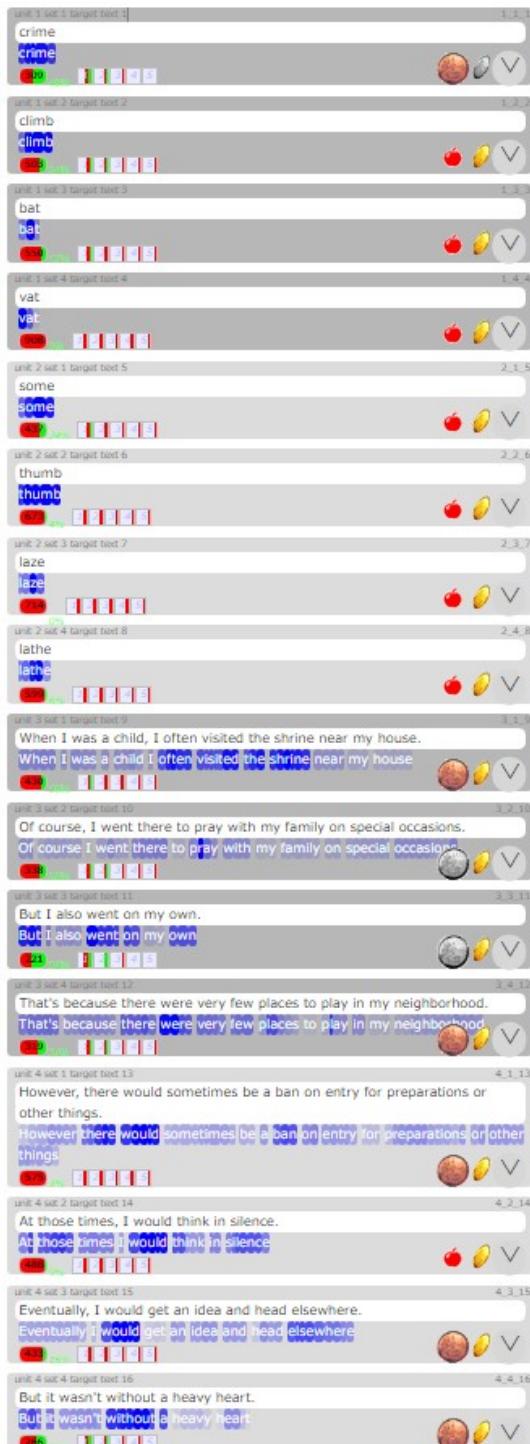
これらの課題を解決するには該当する分野の技術と専門家及び指導者（教員）の協力が必要となる。関心のある研究者や教育者及びその団体の参加を強く期待している。

## 参考文献

1. 今井新悟（筑波大学）, 赤木彌生（山口大学）, 石塚賢吉（株式会社ドワンゴ）, 伊東祐郎（東京外国语大学）, 菊地賢一（東邦大学）, 篠崎隆宏（東京工業大学）, 中園博美（島根大学）, 中村洋一（清泉女学院短期大学）, 西村隆一（和歌山大学）, 本田明子（立命館アジア太平洋大学）, 家根橋伸子（東亜大学）, 山田武志（筑波大学）. 「自動採点スピーキングテスト SJ-CAT の能力推定の検証」  
言語処理学会 第 23 回年次大会 発表論文集  
(2017 年 3 月), 835-838.
2. 中西のりこ, タム・ショウイン, 海老原由貴 (2019). 「ICT 教室の特性と音声認識ソフトを活かした英語リスニング・スピーキング活動の可視化」.  
私立大学情報教育協会 2019 年度 ICT 利用による教育改善研究発表会 [http://juce.jp/archives/ronbun\\_2019/02.pdf](http://juce.jp/archives/ronbun_2019/02.pdf)
3. Spring, R. (2020). Using Multimedia Tools to Objectively Rate the Pronunciation of L1 Japanese ELF Learners. ATEM Journal: Teaching English through Multimedia, 25, 113-124.
4. 井上雄介, 桃島優, 斎藤大輔, 峯松信明 (2018). 「母語話者シャドーイングに基づく可解性自動計測と回帰分析による高精度化」 情報処理学会 第 125 回 SLP 研究発表会. [https://www.gavo.t.u-tokyo.ac.jp/~mine/CALL/ILS/Inoue\\_asj.pdf](https://www.gavo.t.u-tokyo.ac.jp/~mine/CALL/ILS/Inoue_asj.pdf)
5. 湯舟英一, 田淵龍二 (2013). 「映画音声コーパスを利用した Breath Group 長の分析」 Language Education & Technology, 50, 23–41.: [https://www.mintap.com/news/pic/2013\\_7\\_bg.pdf](https://www.mintap.com/news/pic/2013_7_bg.pdf).
6. 田淵龍二 (2018). 「テキストアナリティクスと音声解析と認知科学と検索エンジン」. 電子情報通信学会 言語理解とコミュニケーション研究会 第 13 回 テキストアナリティクス・シンポジウム. [https://www.mintap.com/news/pic/2018\\_nlc13.pdf](https://www.mintap.com/news/pic/2018_nlc13.pdf)

## 付録

### A テスト履歴俯瞰図と課題文全文



1 crime

2 climb

- 3 bat
  - 4 vat
  - 5 some
  - 6 thumb
  - 7 laze
  - 8 lathe
  - 9 When I was a child, I often visited the shrine near my house.
  - 10 Of course, I went there to pray with my family on special occasions.
  - 11 But I also went on my own.
  - 12 That's because there were very few places to play in my neighborhood.
  - 13 However, there would sometimes be a ban on entry for preparations or other things.
  - 14 At those times, I would think in silence.
  - 15 Eventually, I would get an idea and head elsewhere.
  - 16 But it wasn't without a heavy heart.
- (履歴と課題文は5回の授業のうちの初回のもの)

### B NatTos の動作環境など

必須媒体	ネット接続可能な端末
サービス	オープンサイトアプリ/無料無登録
アクセス	<a href="https://www.mintap.com/nattos/">https://www.mintap.com/nattos/</a>
動作確認	Chrome/Android, Windows10
機能	Safari/iPhone, iPad
機能	英語スピーキングテスト, 学習履歴解析
自動採点, 自動赤ペン添削	
テスト課題	<ul style="list-style-type: none"> <li>・ステップアップドリル搭載済</li> <li>・自前文によるテスト作成</li> <li>・作成テスト公開（登録が必要/無料）</li> </ul>
公開	2020年5月
クラス管理	開発中/2021年春公開予定/有料
開発	ミント音声教育研究所（田淵龍二）

### C 受験者（生徒）の動作環境

Chrome	28人
Safari	80人