

日本語学習者の発音能力診断システム開発の基礎研究

畑佐 由紀子 (広島大学)
高橋 恵利子 (目白大学)
山元 啓史 (東京工業大学)
Hodoscek, Bor (大阪大学)
前川 眞一 (東京工業大学)

1. 目的

1.1 はじめに

発音は、母語話者が非母語話者の言語能力を推定するための最も簡易な材料である。文法的に正確に話していても発音が悪いと、能力が低いと評価されたり、発話意図を誤解されたりすることもある。しかし、日本語教育の現場では、発音指導に多くの授業時間を割くことが困難である場合が多い。そのため、外国語として日本語をある程度学習し、来日後に初めて発音の問題に直面し、ストレスを感じる者も少なくない。

グローバル人材交流が拡大する中で、日本語学習者にとって重要なのは、単に通じるだけでなく、適切で好ましい話し方をすることである。学習者自身にとっても、教室外で一般の母語話者に自身の発音がどう評価されるかは、大きな関心事であると考えられる。学習者の日本語発音が一般母語話者にどのように評価されるのかを的確に診断できるような評価システムがあれば、自身の発音能力の把握はもとより、より効率的な発音学習支援が可能となるはずだが、そのようなテストセットや評価システムはいまだ開発されていない。本研究は、日本語学習者の発音の評価システムを開発することを最終目的とするものである。本発表はその基礎研究として、母語話者による学習者の発音評価の方法と、評価傾向について報告する。

1.2 発音評価研究

従来の発音評価研究では、評価者が学習者の発話音声を聞き、項目ごと、あるいは全体的な印象について、4~7段階の尺度で評価する手法が主に取られているが、この方法では、同じ音声であっても評価者の属性や文脈によって評価値が異なることが指摘されている(小池 1998: 151, 小河原 1993: 11, 河野・松崎 1998: 25, 渡辺・松崎 2014: 65, 松崎 2007: 303)。一方で、評価者が一般母語話者であるか音声の専門家であるかに関わらず、ほぼ同様の評価値になるという報告もあり(Schmid and Hopp 2014: 377; Warren et al. 2009: 94)。評価者の訓練なしに尺度法を用いた場合の信頼性については結果が一致していない。また、この方法では尺度や段階の解釈が、評価者間で一致しないことや、中心化傾向(中央を選ぶ傾向)、寛大化傾向(何でもよい方を選ぶ傾向)、ハロー効果(他の要素の影響をうけてポジティブにもネガティブにも歪められる傾向)などが起こりやすい欠点が指摘されている(東他 1973: 444-5)。

こうした問題の解決策として、音響分析を基にして自動的に発音を評価する手法も考えられているが、この手法は評価者の影響を受けないものの、音響分析から算出された音声の逸脱は人間が感じる逸脱と必ずしも一致しないという問題がある。

学習者にとっては、評価者が誰であれ、一般的にどう評価されるのかということが問題となるため、より統合的な評価を還元する必要がある。そこで本研究では、評価者の属性や環境、解釈などの影響を抑えうる、より厳格な評価方法として、一対比較法を採用する。一対比較法は、呈示される2刺激からより好ましい方を選択させる評価方法である。この手法は、類似度の高い刺激でも、その差異を詳しく評価分析することが可能なため、信頼性・妥当性が高い。その一方で、一対比較法は、刺激数が多いと評価作業に時間がかかるというデメリットがある。高橋他(2015)は、この評価方法を用い、6名の学習者の短文音読音声属性の異なる5名の母語話者に評価させた。分析の結果、評価者間評定値に高い相関が認められた。これにより、母語話者であれば、属性に関わらず、安定した評価が得られることが確認できた。

この方法は、厳格で安定した評価が得られる点で優れているが、全ての刺激をペアで比較する必要があるため、参加者や刺激が増えると評価者の負担が増大するというデメリットがある。そこで本調査では、クラウドソーシングを取り入れることにより個々の評価者の作業負担の軽減を図るとともに、人の集合知を利用した評価システムの可能性について検討する。クラウドソーシングは、専門家・非専門家を問わず不特定多数の人々に作業への介入を促し、課題を処理するシステムである。作業をネット上で公開・共有することで、多種多様な人々の集合知を集積することができる(永田 2014: 475, 永田 2013: 373)。本研究の最終的な目的は、学習者の発音に対する一般母語話者の評価をフィードバックすることである。それは不特定多数の母語話者の集合知によって得られるものであり、クラウドソーシングの原理にかなうものである。

2. 調査計画

2.1 データ

データは、米国の夏期集中講座で日本語を学習する初級後半から上級英語母語話者30名(男女各15名)と、母語話者6名(男女各3名)が読み上げた短文40文の音声である。刺激文は、5~24モーラからなる平易な短文で、英語母語話者に指摘される発音上の問題を含むものであった。各音声提供者には、パソコン画面上に読み仮名と英語訳付きの刺激文を視覚提示し、時間制限を設けずに読み上げてもらい、これを録音した。明らかな誤読や録音状態の悪いものを排除し、1417データを評価対象とした。

2.2 手続き

クラウドソーシングによるオンライン評価を実現するため、評価システムを構築し、ネット上にアップロードした(<http://cuckoo.js.ila.titech.ac.jp/app>)。このシステムでは、あらかじめ採取した読み上げ音声をサーバーに保管しておき、評価する音声ファイルを2つずつその都度自動抽出する。ただし、基準音声となる母語話者間の比較は必要ないため、母語話者同士のペアは生じないように設定した。評価者は、評価サイトにアクセスし、評価者登録をする。評価画面ログイン後、システムが自動抽出した2名の読み上げ文の音声を聞いて、A・Bどちらの発音が「日本語としてきれいか」を強制選択する(図1参照)。ブラウザ画面上には読み上げ文の文字情報と選択用のアイコンが示されており、評価者がA・Bいずれかのアイコンを選択(画面をクリックまたはタッチ)すると次のペアが提示される。オンライン上でアクセス可能であるため、評価者は時と場所を選ばず評価に参加できるというメリットがある(図2参照)。一対比較法は、安定した結果が得られる評価方法であるが、相当数の評価回数が必要となる。本稿では、発表グループ内でのパイロット評価の結果の途中経過について報告する。



図1 発音評価システムの web 画面

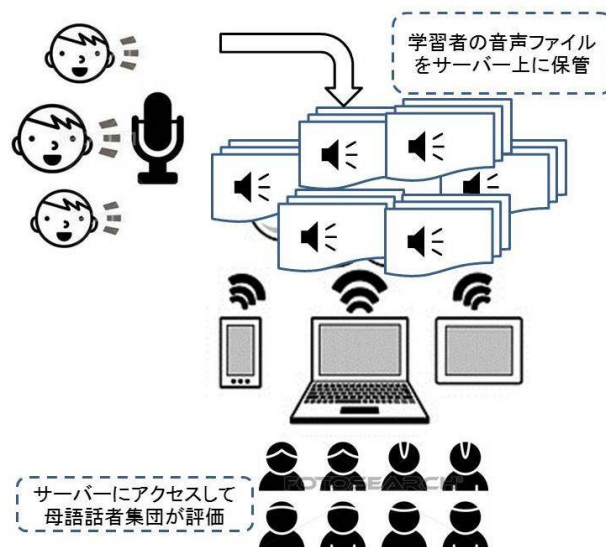


図2 クラウドソーシングの仕組み

3. 分析と考察

調査対象は任意に作成した短文 40 文であるが、システム上では 5 文 1 セットのローテーションで反復提示した。今回の分析対象は最初の 5 文（「こんにちは」「二つで一万円でした」「一緒に旅行に行きましょう」「ここでは写真を撮らないでください」「冷蔵庫に牛乳が入っています」）である。現時点（8 月 8 日）で、5 文合計の評価数は 847、各文の評価数は 170 程度、各話者は母語話者を含めそれぞれ 9～11 回程度、評価されている。9～11 回の評価のうち、よりよいと評価された割合を被選択率として話者ごとに算出した。割合が高いほど、発音がよいと評価されたことを示す。6 名の母語話者は一部の文を除いて、ほぼ 100% 選択されている。しかし、学習者の発音の被選択率は最も高いものでも 60% 程度に止まっており、平均すると 40% 前後であった（表 1 参照）。特に被選択率の低い音声は、発話時間が長い、明らかに発音を誤っている（例、「旅行」→/りっこー/）、など、共通する傾向が見られた。

表 1 より良いと評価された回数の割合（5 文）

	こんにちは	二つで一万円	一緒に旅行に	ここでは写真	冷蔵庫に
母語話者平均	0.83	0.98	1.00	1.00	1.00
学習者平均	0.43	0.39	0.39	0.39	0.38
全体	0.50	0.49	0.49	0.50	0.50

学習者の被選択率が 60% を超えないという結果から、学習者間で発音の差は顕著ではない、ばらつきが少ないということが指摘できる。ただし、評価された回数は現段階では 10 回程度に止まっている。言い換えると 35 名中、最大 10 人の相手としか比較をしていないことになる。今後、さらに評価回数が増えることで差は顕著になることが予想される。また、5 文に対する評価総数 847 回のうち 300 回は母語話者との比較であり、全体の 35% を占める。対母語話者比較の占める割合が多かったことも、学習者の被選択率が伸びなかった原因の一つと推測される。

表 1 を見ると、「こんにちは」に関しては、母語話者対学習者の比較で、学習者の音声を選択されたケースがあったことがわかる。「こんにちは」は短いうえに、学習者にとっても言い慣れ

た言葉である。イントネーションが不自然なものや、撥音が脱落するなどの誤用もあったが、母語話者との差が感じられない音声もあった。こうしたことから、母語話者の被選択率が下がったものと考えられる。

4. 今後の課題

まず最優先すべき課題は、評価回数を増やすことである。評価回数が増えることで、学習者間比較のばらつきがより顕著になると考えられる。それと並行して一対比較の評価データを、多次元尺度構成法の一つである mdpref (Multidimensional Preference Scaling) の最尤解 (Okubo and Mayekawa 2015) で計算する。これにより、不特定多数の母語話者が学習者の発音を評価する際の観点を、多元的に把握することが可能になる。

これと並行して、学習者音声の分析も必要となる。評価に関わるのは、音やリズムの誤りだけでないことは言うまでもないが、どのような発音でどのような要素が優先的に評価に関わるのかといった問題は、十分に探求されていない。分析結果と突き合わせることで、評価に関わる音声的特徴の抽出と特定を進める必要がある。

【参考文献】

- 東洋・大山正・詫摩武俊・藤永保 (編) (1973) 『心理用語の基礎知識：概念の正確な理解と整理』 有斐閣。
- 小河原義朗 (1993) 「外国人の日本語の発音に対する日本人の評価」『東北大学文学部日本語学科論集』, No.3, pp.1-12.
- 小池真理 (1998) 「学習者の会話能力に対する評価に見られる日本語教師と一般日本人のずれ：初級学習者の到達度試験のロールプレイに対する評価」『北海道大学留学生センター紀要』, No.2, pp.138-156.
- 河野俊之・松崎寛 (1998) 「一般日本人と日本語教師の音声評価の差異」『日本語教育方法研究会誌』, Vol.5, No.2, pp.24-25.
- 高橋恵利子・畑佐由紀子・山元啓史・前川眞一・畑佐一味 (2015) 「日本語学習者発音能力評価システムの開発」『人文科学とコンピュータシンポジウム論文集』, No.2, pp.59-64.
- 永崎研宣 (2013) 「人文学分野とサイバーインフラストラクチャー：デジタル・ヒューマニティーズにおける現状と課題」『情報の科学と技術』, Vol.63, No.9, pp.369-376.
- 永崎研宣 (2014) 「日本語クラウドソーシング翻刻に向けて：特集デジタル時代の日本語」『情報の科学と技術』, Vol.64, No.11, pp.475-480.
- 松崎寛 (2007) 「発音評価研究に関する覚書」『大学における日本語教育の構築と展開』 ひつじ書房。
- 渡辺裕美・松崎寛 (2014) 「発音評価の相違：日本人教師・ロシア人教師・一般日本人の比較」『日本語教育』 No.159, pp.61-75.
- Okubo, Tomoya and Shin-Ichi Mayekawa (2015) Modeling viewpoint shifts in probabilistic choice, *Psychometrika*, Vol.80, No.2, pp.412-427.
- Schmid, Monika. S. and Holger Hopp(2014) Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects, *Language Testing*, Vol.31, pp.367-388.
- Warren, Paul, Irina Elgort, and David Crabbe(1984) Comprehensibility and prosody ratings for pronunciation software development, *Language Learning and Technology*, Vol.13, pp.87-102.