

外国人日本語学習者の発音能力を測定するシステムの開発

高橋 恵利子¹ 畑佐 由紀子² 山元 啓史³ 前川 眞一³ 畑佐 一味⁴

概要：本研究の目的は、外国人日本語学習者の発音能力を簡易に診断するシステムを開発することである。その目的を達成するためにはさまざまな問題があるが、本稿では音声データの収集形式の問題と評価者の問題を取り上げる。音声データの収集方法としては、短文を読み上げ、それを録音する方法（読み上げ課題）と、同じ短文をあらかじめ録音したものを聞いて発音したものを録音する方法（リピート課題）の2つを検討する。録音の評価者は、全員日本語母語話者（日本語教師、日本語教育未経験者）とし、これらの条件で、6名（母語話者2名、ほぼネイティブ水準の発音技能を持つ者2名、顕著な外国人訛りを持つ者2名）の音声提供者の録音資料を用い、一対比較法による評価実験を行った。実験の結果、データの収集方法については、いずれの方法によっても0.86以上の相関係数が得られたが、リピート課題（0.86以上）よりも読み上げ課題（0.92以上）の方が、若干高かった。これにより、今後のシステム設計計画では、一般の母語話者を評価者とし、あらかじめ音声材料を準備する必要のない読み上げ課題によるデータ収集方式を採用することにした。

キーワード：日本語教育、音声の指導、診断、一対比較法、MDPREF、評価者間比較

1. はじめに

外国語の発音指導に付随して、発音をいかに評価するかという問題がある。発音に特化した評価システムが開発できれば、学習者は自身の発音上の問題を把握することができ、改善点を明確化、意識化することで自律的な学習につなげることができる。これまでの日本語教育における発音評価の主流は、評価者が学習者の発話音声聞き、項目ごと、あるいは全体的な印象について、数段階の尺度で評価するもので、評価の決定要因を探索的に特定することが主目的だった。従来の研究では、同じ音声であっても評価者によって、また文脈や状況によって、評価値が異なることが指摘されている。近年、評価者の影響を受けない自動評価システムの開発が進められているが、言語学習に還元す

るためには、単に規範からの逸脱を指摘するだけでなく、逸脱のタイプや程度と評価値との関係を明らかにする必要がある。

本稿では、システム開発の事前調査として評価者の均質性を検証することを目的とする。これまでの発音評価研究では、評価者の属性差に着目したものが多く、日本語教師と一般母語話者（[1]: 151; [2]: 183; [3]: 25）、また母語話者教師と非母語話者教師（[4]: 13; [5]: 65）とでは、評価の観点が異なること、外国人との接触度が評価に影響すること（[6]: 29）などが報告されている。また、評価者の母方言が統制されることも多い。しかし、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。これまでの評価研究は、評価者の属性による評価の差が留意事項になりがちであったが、学習者にはより統合的な評価を還元する必要がある。

現実的な評価方法として、具体的にどのような方式を採用するかも課題である。従来の評価研究では4-7段階の評定尺度法（[7]: 105-6）が一般的に用いられているが、この方法では尺度の目盛あるいは段階の解釈が評価者の独自の解釈に依存してしまうことや、中心化傾向（中央を選ぶ傾向）、寛大化傾向（何でもかんでもよい方を選ぶ傾向）、ハロー効果（対象者に期待をするあまりそれに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（[8]: 444-5）。また、直

¹ 目白大学: e.takahashi@mejiro.ac.jp; Mejiro University, Faculty of Foreign Language Studies, 4-31-1, Naka-Ochiai, Shinjuku, Tokyo 161-8539, Japan.

² 広島大学: yhatasa@hiroshima-u.ac.jp; Hiroshima University, Graduate School of Education, 1-1-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8524 Japan.

³ 東京工業大学: {yamagen@ryu,mayekawa@hum}.titech.ac.jp; Tokyo Institute of Technology, Department of Human System Science, Graduate School of Decision Science and Technology, 2-12-1 O-okayama, Meguro, Tokyo 152-8550, Japan.

⁴ パデュー大学: khatasa@purdue.edu; Purdue University, Department of East Asian Languages School of Languages and Cultures 640 Oval Drive West Lafayette, IN 47907-2017, USA.

表 1 読み上げ課題, リピート課題に用いられた文のリスト; SID は
センテンス ID.

SID	課題文
S01	ここが玄関です。
S02	亡くなったおばあちゃんの写真です。
S03	大盛りです。
S04	冷凍食品はアメリカの方が多いですね。
S05	1200円になります。
S06	これは冷蔵庫です。
S07	息子さんはおいくつですか。
S08	これが一番使いやすいですね。
S09	1万円お預かりいたします。
S10	中は何になりますか。

前刺激との相対評価になりやすいため, 提示順序によって
評価値が変わる可能性も排除できない。そこで, 本研究で
は評価すべき音声を総当たりで組み合わせ, それらを一対
ごとの比較で評価を試み, 評価者間の評価傾向に違いが見
られるかどうかを検討する。

2. 方法

2.1 材料

発音課題用の短文として, 比較的自然的な状態で録音され
た材料(オーセンティックな材料)を使用することとし,
テレビ番組や実際の会話から抜き出した10の短文を採用
した(表1)。

2.2 音声提供者

音声提供者(いわゆる発音を評価される「学生」に当た
る)は2名の母語話者と4名の中国人日本語学習者であ
る。尺度を網羅したデータを得るために, 最高の技能を持
つ者として母語話者(NS1, NS2), ほぼネイティブ水準の
発音技能を持つ者(CH1, CH2), 顕著な外国人訛りを持つ
者(CM1, CM2)を採用した。

2.3 手続き

課題は読み上げ課題とリピート課題の2種類を用意した。
音声提供者には, コンピュータ画面に呈示される文を読み
上げることを求めた。同時に, 聴覚呈示される文を即座に
リピート再生することも求めた。これは文末イントネー
ションの理解を統制することを目的としている。読み上げ
課題とリピート課題は1文ずつ交互に採取した。なお, 課
題文をパワーポイント(プレゼンテーション用提示ソフト
ウェア, PPT)で呈示する際には, 漢字には読み仮名をつ
け, さらに中国語訳も添えて呈示した(図1)。調査に用い
た課題文は表1の通りである。各音声提供者から読み上げ
課題, リピート課題の各10個の音声を得た。

評価用刺激として, 1文ごとに6名の音声を総当たりで
ペアにして, 読み上げ課題の音声, リピート課題の音声, 各

150個(5×6/2)の音声刺激ファイルを作成した。全300
個の刺激をランダムに配置し, 5名の評価者(教師1名,
大学院生2名, 学部生2名)にランダムで聴覚呈示した。
評価者は任意2名の音声提供者による録音を聞き, 直感的
に, より自然だと思った方をキー入力で選び, 回答するこ
とが求められた。評価の所要時間は1人あたり40分前後
であった。なお評価者の母方言は問わないこととした。

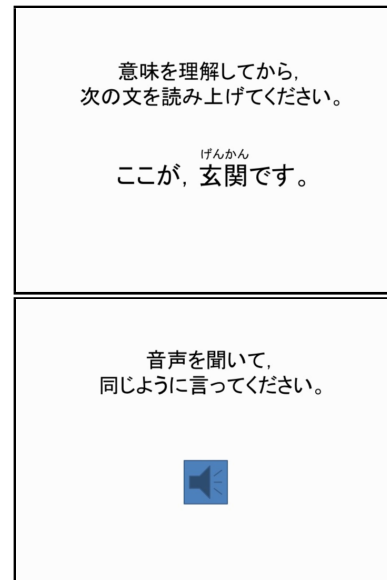


図1 コンピュータによる読み上げ課題(上)とリピート課題(下)の
教示方法

3. 結果

一対比較のための音声を5名の評価者が評価し, その
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[9]で計算
した*1。以下では, その結果を評価者と課題文の観点から
述べる。

3.1 評価者

評価者(E00001..5)と音声提供者(NS1, NS2, CH1, CH2,
CM1, CM2)の関係をbiplotで表したのが, 図2(読み上
げ課題), 図3(リピート課題)である。

読み上げ課題については, すべての評価者同士の相関は
0.922以上であった。リピート課題についても, E00001と
E00004の評価者間の相関が, 0.861である他は, いずれも
0.936以上の相関があった。MDPREFのd1軸の相対的な
寄与は, 読み上げ課題で0.970, リピート課題で0.962で
あり, いずれにおいても一元的な尺度と見てよい。

*1 MDPREF用のRパッケージは, 前川研究室のRパッケージの
ページ(<http://www.ms.hum.titech.ac.jp/Rpackages.html>)
から利用できる。

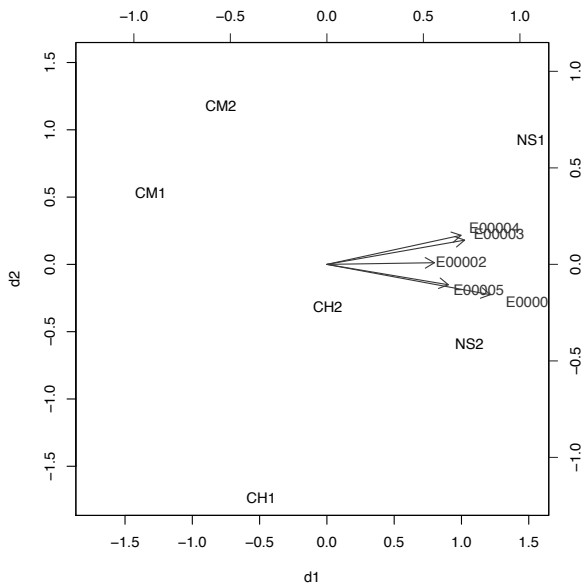


図 2 読み上げ課題における評価者 5 名 (E00001..5) と音声提供者 6 名 (NS1, NS2, CH1, CH2, CM1, CM2) の biplot.

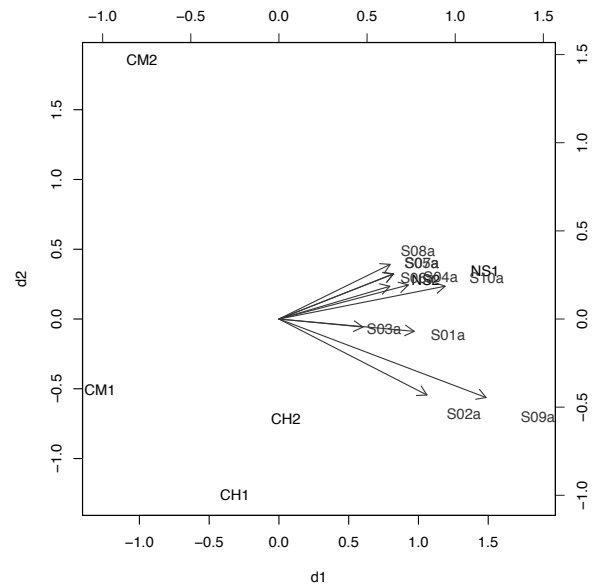


図 4 読み上げ課題における課題文 10 (S01a..S10a) と 6 名の音声提供者 (NS1, NS2, CH1, CH2, CM1, CM2) の biplot.

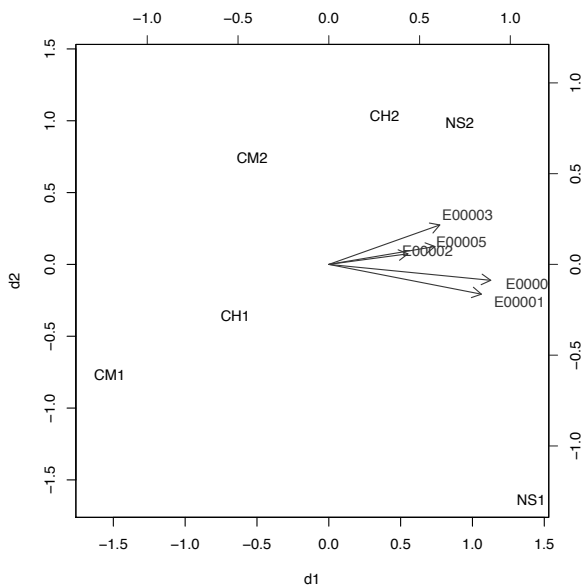


図 3 リピート課題における評価者 5 名 (E00001..5) と音声提供者 6 名 (NS1, NS2, CH1, CH2, CM1, CM2) の biplot.

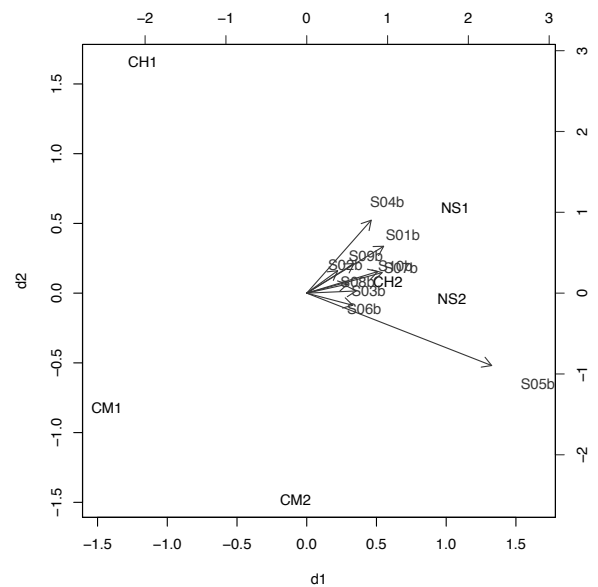


図 5 リピート課題における課題文 10 (S01b..S10b) と 6 名の音声提供者 (NS1, NS2, CH1, CH2, CM1, CM2) の biplot.

3.2 課題文

課題文 (S01..S10) と音声提供者 (NS1, NS2, CH1, CH2, CM1, CM2) の関係を biplot で表したのが、図 4 (読み上げ課題)、図 5 (リピート課題) である。読み上げ課題、リピート課題のいずれにおいても、センテンス毎の課題のベクトルは右向きを示しており、MDPREF の d1 軸の相対的な寄与は、読み上げ課題 0.892、リピート課題 0.917 であった。ただし、読み上げ課題とリピート課題とでは、CH1 の布置が大きく変わり、読み上げ課題では、NS1-NS2-CH2-CH1-CM2-CM1 の順に、リピート課題では、

NS1-NS2-CH2-CM2-CH1-CM1 の順になった。

読み上げ課題で顕著なベクトルは課題文 S09a である。S09a に類似する課題文は S02a であり、相関係数 0.994 であったが、音声特徴上の共通点は両者には見られない。S09a に直交する軸で音声提供者の布置を見ると、NS1-NS2-CH2-CH1-CM2-CM1 の順となった。これは、ネイティブ (NS) → ネイティブ水準 (CH) → 外国人訛りあり (CM) の順序と一致し、評価としては妥当である。

リピート課題で顕著な課題文は S05b であり、それと最も相関が高いのは、S06b (0.993) であった。逆に、S05b

と最も相関が低いベクトルは S04b (0.344) であり、むしろ、S05b よりも S04b の方が他の課題文との相関は高かった (たとえば、S04b と S09b は 0.959, S01b は 0.955, S02b は 0.976 など)。

課題文 S05b と S04b で直交する場合の音声提供者の配置を見てみると、S05b に直交する軸では、NS2-NS1-CM2-CH2-CM1-CH1 の順となった。CM2 がネイティブ (NS) に迫る高評価である一方、CH1 が外国人訛りあり (CM) よりもずっと低い評価となっている。S04b に直交する軸では、NS1-NS2-CH1-CH2-CM2-CH2 の順となり、ネイティブ (NS) → ネイティブ水準 (CH) → 外国人訛りあり (CM) の順序と一致する。

4. 考察

4.1 評価者と評価方法

実験の結果、読み上げ課題とリピート課題との間のそれぞれの評価において、大きな差は見られなかった。したがって、読み上げ課題によって得られた音声だけでも評価が可能であることがわかった。本来、評価されるべき技能は「学習者自身の持つ自然な発音」であるから、読み上げ課題によって評価できることは目的に即している。

つぎに、5名の評価者に高い相関が認められたことから、評価者を日本語母語話者に限れば、ほぼ同じような結果が得られることもわかった。これは、どちらが良いか、必ず二者択一を迫る対比較法を用いたためと考えられる。しかし、いつでもこの方法が実施できるわけではない。この点については後述する。

本研究では、学習者の発音誤用に関する分析は行っていない。刺激文によっては5名の評価がわかる文があった。それらについては、今後、評価に影響を与える発音上の特徴について、記述的かつ音響的な分析を行い、誤用の特性と評価値との関係を探る必要がある。

課題文は「現実の発話であること (authenticity)」を重視したが、文によっては学習者の発音上の問題が現れにくいものもあった。学習者の母語やレベルをある程度絞ったうえで、問題が現れそうな特徴を取り入れた文を用意する必要がある。また、評価者の負担を考慮して課題文の数を10に限定したが、発音誤用を網羅的に検査するには、負担のない方法を採用し、かつ多様な視点を持つ課題文のセット数を検討する必要がある。評価者の均質性が保証できれば、大量のセットを複数の評価者で分担評価し、結果を統合することも可能である。

4.2 評価システムの設計

対比較法 ([8]: 445, 507) は、本研究の実験でわかるように、厳格で安定した評価値が得られる点で優れている。しかし、総当たりペアの比較が条件なので、課題文と評価すべき学生の数が増えると、比較すべきペアの数が激増して

しまう。そこで、本研究を通して暫定的ではあるが、これまでに得られた尺度を更新する形で、極限法 ([10]: 158) の中の上下法 (up-and-down method) を組み合わせるなどして、自動的に自明なペアの判断を省略する仕組みが必要である。

5. おわりに

本研究では、評価者の嗜好の影響を受けない自動評価システムの開発のために、複数の音声提供者による録音資料を総当たりで組み合わせ、対比較という必ずどちらか一方を選ばなければならない方法を用い、評価者間で評価傾向に違いが見られるかどうかを検討した。その結果、読み上げ課題、リピート課題のいずれにおいても、また、いずれの評価者間においても、評価に影響がほとんど見られないことがわかった。

参考文献

- [1] 小池真理：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, Vol. 2, pp. 138–156 (1998).
- [2] 中川道子, 石島満沙子：会話の上達度を計る評価基準, 北海道大学留学生センター紀要, Vol. 2, pp. 169–185 (1998).
- [3] 河野俊之, 松崎 寛：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 (1998).
- [4] 渡部倫子：日本語口頭運用能力の評価基準—評価者による相違—, 日本教科教育学会誌, Vol. 25, No. 4, pp. 11–17 (2003).
- [5] 渡辺裕美, 松崎 寛：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, Vol. 159, pp. 61–75 (2014).
- [6] 小河原義朗：日本語非母語話者の話す日本語の発音に対する日本人の評価意識, 日本語教育方法研究会誌, Vol. 8, No. 1, pp. 28–29 (2001).
- [7] 芝 祐順, 渡部 洋, 石塚智一：統計用語辞典, 新曜社 (1984).
- [8] 東 洋, 大山 正, 詫摩武俊, 藤永 保 (編)：心理用語の基礎知識: 概念の正確な理解と整理, 有斐閣 (1973).
- [9] Okubo, T. and Mayekawa, S.-I.: Modeling viewpoint shifts in probabilistic choice, *Psychometrika*, Vol. 80, No. 2, pp. 412–27 (2015).
- [10] 市川伸一：心理測定法への招待: 測定からみた心理学入門, サイエンス社 (1991).