

日本語単語の韻律モーフィング音声に対する母語話者と学習者による 自然性評価の比較*

◎加藤集平[†], ショート グレグ[‡], 峯松信明[†], 鶴谷千春[‡], 広瀬啓吉[†] ([†]東大, [‡]グリフィス大)

1 はじめに

外国語学習者が自然な発音をするためには、学習者の母語 (L1) と対象言語 (L2) の音声音響学的差異を意識し、それに基づいて学習戦略を立てることが重要である。しかし母語干渉のために、これはしばしば困難な作業となる。特に、韻律は学習者にとって意識的に矯正することが難しい [1]。そのため、まず、L1 と L2 の間の韻律的差異に気付かせることが重要である。ところが留学生を対象とした大学の日本語教育では、時間的制約から、学習者が発音、特に韻律を学ぶ機会が少ない。多くの学習者が日本語のピッチアクセントの知識を持っていないのである。このような状況では、学習者自身が自らの発音の不自然性に気付くのは非常に難しい。近年では、日本語教育における韻律の重要性が目立つようになっている [2]。

母語干渉は、対象言語を母語とする話者には、不自然な発音 (訛) として知覚されることから、どの種類の訛が自然性を最も劣化させるのかを調べる必要がある。そのためには、様々な種類、程度の外国語訛が含まれる音声サンプルの自然性を、母語話者に主観評価させる必要がある。[3-5] では、非母語話者によって発声された発音を母語話者に評価させ、明瞭性や自然性の評価を行っている。これらの研究では、学習者の発声そのものを音声刺激として用いている。しかしこの方法では、幅広い種類の訛を網羅したり、異なる音響特徴量を独立に扱うことが難しい。一種類の訛のみに注目する場合も、異なる訛の程度で音声サンプルを集めるのは困難であろう。そこで必要となるのは、様々な種類、程度の外国語訛が含まれる音声サンプルを用意する方法である。

方法の一つに、音声モーフィングの使用がある。高精度なモーフィングツールの一つに STRAIGHT [6, 7] があり、年齢、性別、感情、訛のモーフィングなど様々な方面に応用されている。[8] では、right と light のようにミニマルペアをなす 2 つの発音を同一話者に発声させているが、STRAIGHT はその 2 つの発音の中間の発音を合成できる。right と light の間で、任意のモーフ率で定量的な補間を行えるのである。

筆者らは以前、[8] と同様に、同一話者による日本語単語の発声対 (自然な発音と訛った発音) を用いて中間の音声刺激を合成し、それらの刺激の自然性を日本語母語話者に評価させた [9]。STRAIGHT を用いると、特定の音響特徴量について音声を定量的に

モーフすることで、外国語訛をその音響的側面から、より正確に扱うことができる。[9] では、実験の結果、音響特徴量や L1 によって、日本人聴取者が評価する自然性にどの程度影響するかが様々であることを示し、どの種類の訛が最も自然性を劣化させるのかを明らかにした。

しかしながら、母語干渉による発音の問題を直すためには、学習者が気付いていない音響の特徴を、学習者に知らせることが必要である。もし彼らが、(母語話者にとって) 何が不自然であるかに気付いていなければ、日本語を自然に話すことは難しいであろう。そこで本論文では、豪人日本語学習者に [9] で用いた音声刺激の一部を聴取させ、[9] と同様に自然性の評価を行わせた。そして、日本語母語話者と学習者による評価様式の違いを比較した。

2 実験

2.1 音声収録

日本語と他の 1 言語を母語話者程度に話せると社会的に認識されている、発話者 2 名の日本語単語発声を収録した。発話者の言語背景を Table 1 に示す。

単語はアクセント型、モーラ数、特殊拍の有無及び単語内の位置についてのバランスセットとなるように選んだ。総単語数は 162 (名詞 112, 動詞 30, 形容詞 20) であった。

発話者には各単語を、東京方言と、(その人が話す他の 1 言語の) 外国語訛を想定した発音で発声させた。同一話者の発声対を用いるのはやや不自然であるが、これは同一話者間のモーフィングの方が、異なる話者間より高品質の合成音声を得られるためである。東京方言での発声では、発話者は読み上げリストとして全ての単語が平仮名 (単語アクセント (NHK アクセント辞典に準拠 [10]) 表示あり) 及び漢字で表記されたものを用いた。名詞はキャリアセンテンス (これは、[名詞] です。) 中で発声させた。外国語訛を想定した発声では、発話者に、読み上げリストの単語表記として平仮名 (単語アクセント表示なし)、ローマ字 (ヘボン式)、ローマ字 (米語の母語干渉を反映したもの。第二著者 (米語母語話者) が作成)、ハングル (標準表記法に準拠 [11]) のいずれを用いるかを選ばせた。韻律的な母語干渉の影響に注目できるように、発声の際にはできるだけ音素は正しく (東京方言と同様に) 発声するように依頼した。

* Comparison of native and non-native evaluations of the naturalness of Japanese isolated words with prosody modified through voice morphing. by KATO, Shuhei, SHORT, Greg, MINEMATSU, Nobuaki (Univ. of Tokyo), TSURUTANI, Chiharu (Griffith Univ.), and HIROSE, Keikichi (Univ. of Tokyo)

Table 1 発話者 2 名の言語背景

性別	年齢	言語	日本以外の居住歴	読み上げ時に使用した表記
女	30 代前半	米語	6～14 歳 米国, カリフォルニア州 (現地の学校に通学)	母語干渉を想定したローマ字表記
女	30 代前半	韓国語	0～23 歳 韓国, ソウル (現地の学校に通学)	平仮名表記

Table 2 モーフィングに使用した音響パラメータ

パラメータ	略称
基本周波数	F0
音素の継続長	dur
スペクトル包絡及び非周期性指標	sp_ap
基本周波数及び音素の継続長	F0_dur
全て	all

2.2 モーフィング

音声刺激は、各話者ごとに、東京方言と訛った発音の単語発声対をモーフすることで合成した。STRAIGHT では、どの音響特徴量をモーフするかを選べる。つまり、4 音響特徴量（基本周波数、音素の継続長、スペクトル包絡、非周期性指標）を独立にモーフできる。スペクトル包絡をモーフすると、スペクトルの形、及びエネルギーが変化する。非周期性指標は音声の有声度を表す。本論文では、Table 2¹に示す 5 音響特徴量と 5 モーフ率 (0, 0.25, 0.5, 0.75, 1) を使った。0 は東京方言での発音を、1 は訛った音声に完全にモーフされていることを意味している。

このようにモーフィングを行うことで、どの音響特徴量を、どの程度モーフすると有意に自然性が劣化するかを調べることができる。モーフィングの例を Fig. 1 に示す。

最終的に、各単語ごとの音声刺激の数は 42 (= (1+4 モーフィング率 × 5 音響特徴量) × 2 言語) となった。単語数は 162 であったので、聴取実験に使用するサンプルの総数は 6,804 となった。

2.3 被験者

2.3.1 日本語母語話者

日本語母語話者 42 名 (19～28 歳、多くは大学生) が聴取実験に参加した。音声刺激は 4 つのサブセットに分割し、各サブセットを 10～12 名の被験者に評価させた。各サブセットは更に複数のセッションに分割して実験を行い、被験者にはセッション間で休憩を取ることを許した。実験は Web を通じて行った。被験者はヘッドホンを通じて音声刺激を聴取し、画面上のボタンをクリックして自然性の評価を入力した。

なお、聴取実験の前に、被験者に 7 段階の自然性と音声刺激との関係をつかんでもらえるよう、15 のサンプル刺激を示した。

¹sp と ap は独立にではなく、常に同時にモーフした。独立にモーフするとしばしば不自然な (人間の発声らしくない) 音になる場合があったためである。考察の節では、sp_ap を一つの音響特徴量として考える。

2.3.2 豪人日本語学習者

豪人日本語学習者 15 名 (18～30 歳、大学で日本語を学んでいる学生) が聴取実験に参加した。彼らは大学での授業を含め、3～10 年間日本語を学んでいる。但し、日本語母語話者を対象とした実験で使用した単語の多くは豪人学習者には難しすぎると考えられたため、使用する語彙は日本語能力試験 2 級以下に出題されるものに限定した [12]。その結果、単語数は 20、音声刺激の総数は 840 となった。音声刺激を 2 つのサブセットに分割し、各サブセットを 7～8 名の被験者に評価させた。他の条件は日本語母語話者に対する実験と同様とした。

2.4 タスク

2 種類のタスクを被験者に課した。1 つ目のタスクは音声刺激の日本語 (東京方言) としての自然性を、1 (非常に不自然) から 7 (非常に自然=母語話者らしい) の 7 段階で評価することである。刺激音声はランダム順に示された。豪人学習者に対しては、日本語母語話者に近い条件で自然性を評価できるように、聴取している音声刺激の発音を平仮名 (アクセント表示なし) で画面上に表示し、それとは別に平仮名 (アクセント表示あり) で表記した単語リストを用意した。2 つ目のタスクは、音声刺激が非常に合成音らしいかどうかを判定することであった (STRAIGHT でモーフィングを行うと、人間の発声らしくない音声合成される場合がある)。被験者には音声刺激が合成音であることを予め伝えていたが、非常に合成音らしいと判定された音声刺激に対する自然性評価は信頼性に乏しいと考えられる。以後の節の議論では、過半数の被験者に非常に合成音らしいと判定された刺激は、考察対象外とした。

3 結果と考察

3.1 聴取実験の結果

実験結果を Fig. 2 に示す。グラフは上段から英語訛 (日本語母語話者による評価)、同 (豪人による評価)、韓国語訛 (日本語母語話者による評価)、最下段が韓国語訛 (豪人による評価) である。横軸はモーフ率、左縦軸は被験者によって評価された主観的自然性、そして右縦軸は p 値である。p 値は一元配置分散分析によって計算されたものであり、あるモーフ率での p 値は、主観的自然性がモーフ率 0 (自然な発音) のそれと有意に異なると言える、最低の有意水準を

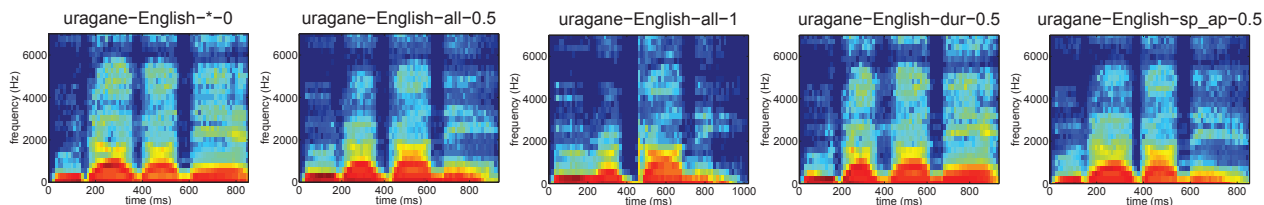


Fig. 1 単語「うらがね」(裏金)をモーフした音声のスペクトログラム

与えるものである。すなわち p 値が小さいほど、そのモーフ率での自然性と東京方言の自然性の差がより大きいことを意味している。

モーフ率、特徴量、言語が与えられた時、主観的自然性は以下のように計算される。

1. 各単語について、自然性の評価を被験者間で平均する。
2. 1. で計算した平均評価を単語間で平均する。標準偏差も計算する。

平均と標準偏差 ($\pm\sigma$) は Fig. 2 に示されている。

3.2 考察

各訛においてどの音響特徴量が自然性に最も影響しているかを調べた。そのために、各グラフにおいて、主観的自然性が有意に劣化する最小のモーフ率を計算した。最小のモーフ率が小さいほど、その特徴量の自然性への影響は大きいと言える。

まず、個々の音響特徴に対する日本語母語話者の結果を調べた。英語訛の刺激に対しては、 p 値 < 0.01 となる最低のモーフ率は 0.5 (F0, dur, sp_ap) である。韓国語訛に対しては、0.5 (F0, dur), 0.75 (sp_ap) である。次に、豪人学習者の結果を調べた。英語訛の刺激に対しては、 p 値 < 0.01 となる最低のモーフ率は 1 (F0), 0.75 (dur), 0.5 (sp_ap) である。韓国語訛に対しては、存在しない (F0, dur), 1 (sp_ap) である。これらの結果から、日本語母語話者と豪人学習者間で主観的自然性が最も異なるのは F0 で、次いで dur となった。一方、sp_ap は両者で似たような自然性のパターンを示した。

これらの結果と、F0, dur, 及び sp_ap のグラフにおける自然性のパターンを観察して言えるのは、各言語訛ごとに、各音響特徴量が主観的自然性に与える影響の程度は異なり、その程度は日本語母語話者と豪人学習者の場合で異なるということである。日本語母語話者による英語訛の評価では、F0, dur, 及び sp_ap は似たような主観的自然性の劣化の様子を示している。しかし豪人学習者による評価では、sp_ap が最も激しく劣化し、dur, そして F0 と続く、韓国語訛の場合も同様の傾向である。

日本語母語話者にとって F0 パターンの変化が自然性に大きな影響を与えている一方、豪人学習者にとっては非常に小さな影響しか与えていない。これは、日本語のピッチアクセントを体系的に学ぶ時間の不足が原因であると思われる。よく知られているように、

多くの日本語教科書はピッチアクセントの表示がない。より多くの時間をピッチアクセントの体系的な学習のために確保し、日本語母語話者が不自然と判断する F0 パターンが、不自然であることに気付かせることが必要であろう。

F0, dur, 及び all, すなわち複数の特徴量の組み合わせの結果は、それぞれを構成する個々の特徴量の影響として解釈することができる。

3.3 教育学的観点からの提案

本研究で明らかになったのは、学習者は継続長誤りは比較的に正確に検出できた（日本語母語話者ほどではないが）一方、ピッチ誤りはできなかったということである。一般に、学習者にとってピッチの逸脱は継続長のそれよりも検出するのが難しく、訓練されていない聴取者はピッチの上昇／下降をしばしば判断できない [13]。加えて、多くの日本語教科書は新出語彙のピッチアクセント表示がなく、ピッチアクセントの上達は学習者自身に放任している状態である。結果として、学習者は正しいピッチパターンを、音声資料や母語話者の友人などを通じて、偶然に学ぶしかなくなる。語学教師は、学習者に学習言語でとりあえずコミュニケーションさせるために、ピッチ誤りを文法や発声内容の適切さに比べて些細な問題として扱う傾向がある。学習者が教師や母語話者と会話をする際に、個々の発音誤りは意思疎通を必ずしも阻害しない。それでも、新出語彙が紹介された時には、より明確にピッチパターンを教えるべきである。日本語教室にとって理想的なツールの一つは、学習者の発声のピッチ曲線を解析し、クイズやゲーム形式で学習者のピッチ誤りを指摘できるコンピュータプログラムであろう。ピッチの視覚的表示は、学習者が自らの発音のピッチの上昇／下降が分からない時に、特に有用であろう。まずは、学習者の注意を重大なピッチ誤りに向けさせることである。自らの発音の音素誤りを検出できるコンピュータプログラムに関しては、学習者が楽しんで使うという報告がある [14]。ピッチ誤りを検出するプログラムの発達に、語学教室で待たれている。

4 おわりに

本論文では、日本語の発声においてどの種類、程度の外国語訛が、日本語母語話者及び豪人日本語学習者が感じる自然性に影響を与えるのか、及び両被験者間の評価の違いを調べた。音声刺激としては、

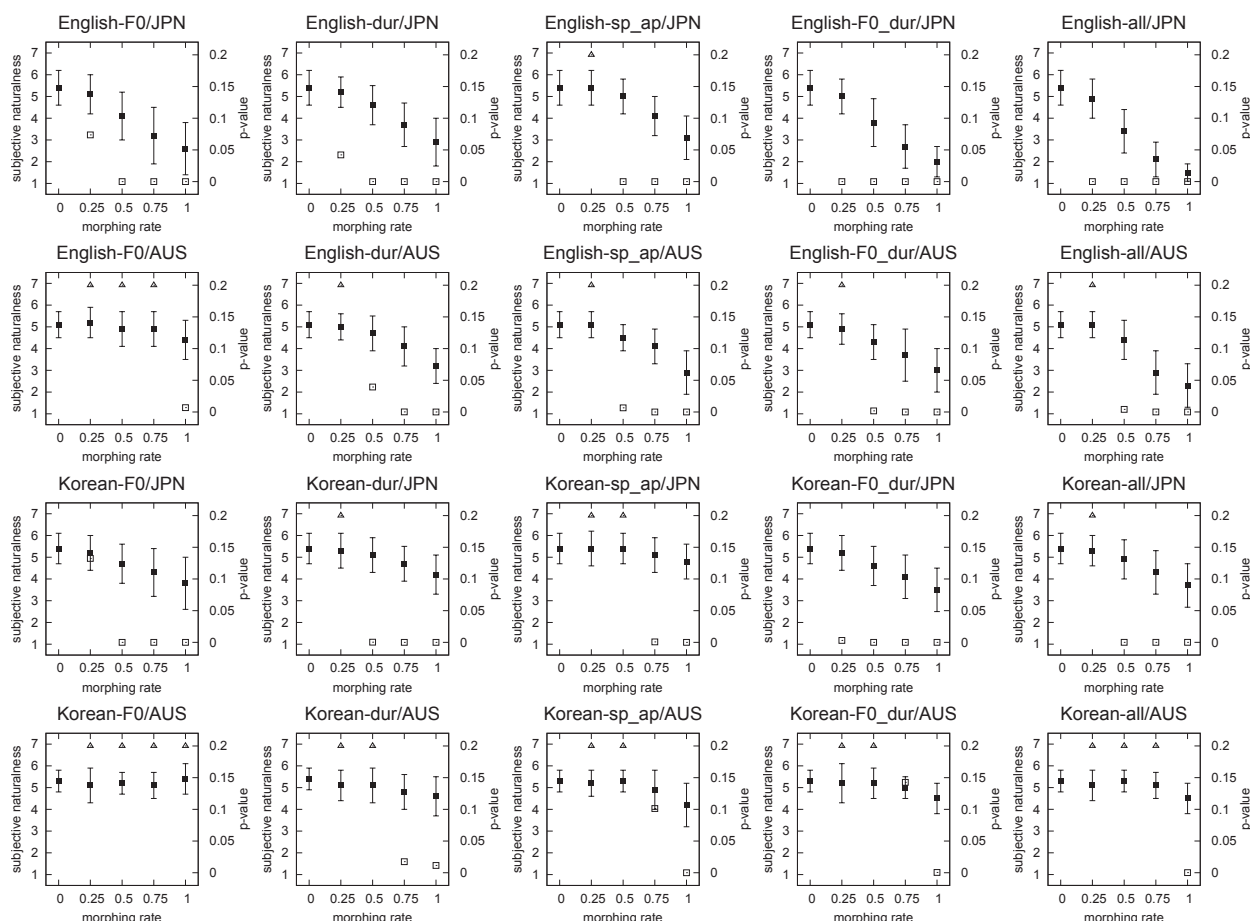


Fig. 2 言語別、特微量別の主観的自然性。■：主観的自然性；□：p 値 (△ は p 値 > 0.2 であることを示す)。

STRAIGHT モーフィングを用いて、様々な種類と程度の英語訛及び韓国語訛の日本語単語発声を用意した。これらの刺激を日本語母語話者と豪人学習者の両聴取者に聴取させ、日本語（東京方言）としての自然性を評価させた。実験結果より、各言語訛、各音響特微量によって母語干渉が発音の自然性に与える影響の程度が異なること、その程度は日本語母語話者と豪人学習者の間で異なることが示された。特に、豪人学習者は F0 パターンの変化に鈍感であった。そこで、なぜ豪人学習者が特定の音響特微量について母語話者と同様に自然性を評価できないのかについて考察した。最後に、本研究が語学教室でどう応用されるかについて、いくつかのアイデアを述べた。

今後は、豪人以外の日本語学習者、例えば中国人や韓国人に対して同様の実験を行うことを検討している。また、本論文の結果は発話者 2 名の発声のみに依存しているため、より多くの発話者を用いることも検討している。

参考文献

[1] T. Shibata *et al.*, “Prosody Acquisition by Japanese learners,” In *Understanding Second Language Process*, Multilingual Matters, 2007.

[2] 中村他, さらに進んだスピーチ・プレゼンのための日本語発音練習帳, ひつじ書房, 2009.
 [3] J. Bernstein, *Proc. ICPhS*, 1581–1584, 2003.
 [4] N. Minematsu *et al.*, “Measurement of objective intelligibility of Japanese accented English using ERJ database,” *Proc. INTERSPEECH*, 2011.
 [5] C. Tsurutani, *Proc. INTERSPEECH*, 1854–1857, 2010.
 [6] H. Kawahara *et al.*, *Speech Communication*, 27(3–4), 187–207, 1999.
 [7] H. Kawahara *et al.*, *Proc. ICASSP*, 256–259, 2003.
 [8] 久保他, 音講論 (春), 1-8-22, 383–384, 1998.
 [9] 加藤他, 信学技報, SP2010-118, 19–24, 2011.
 [10] NHK 放送文化研究所編, “NHK 日本語アクセント辞典 新版,” NHK 出版, 1998.
 [11] 大韓民国文化体育観光部国立国語院, “外来語表記法,” http://www.korean.go.kr/09_new/dic/rule/rule_foreign.0104.jsp
 [12] 国際交流基金, 日本国際教育支援協会, 日本語能力試験, <http://www.jlpt.jp>
 [13] 鶴谷, 第二言語としての日本語の発音とリズム, 溪水社, 2008.
 [14] C. Tsurutani, *Journal of Japan Studies Association of Australia*, 305–315, 2008.