

Praat を用いた日本語イントネーションの知覚研究 (A Perceptual Study of Japanese Intonation Using Praat)

新谷 敬人 (Takahito Shinya)
マサチューセッツ大学アマースト校大学院

1. はじめに

本稿では、音声分析・合成ソフトPraat (Boersma & Weenink 1992-2006)を用いた日本語イントネーションの知覚研究を、その刺激音作成と実験実施の点に重点を置いて解説する。音声知覚の研究では、いかに刺激音の音響特性をコントロールするかが重要なポイントとなるため、合成音の使用が必要になる。しかし、最もよく使われるタイプの合成音は、人間が発話した音声を基にして音響特性を操作し、再合成したものである。ここでは基本周波数(F0)を操作して行ったイントネーションのプロミネンス(prominence)に関する研究の一部を報告する(Shinya 2005, 2006, forthcoming)。

2. 研究の概要

プロミネンスの知覚研究には Gussenhoven ら(1997)に代表される一連の研究がある (Gussenhoven & Rietveld 1988, Gussenhoven et al. 1997, Pierrehumbert 1979, Terken 199)。これらの研究では、オランダ語あるいは英語を対象とし、発話に生じる F0 ピークの値を音声合成により様々に変化させて、プロミネンスの知覚にどのような影響を及ぼすかが検討されている。しかし、これらの研究には1つの重要な視点が欠けている。語彙的トーン(lexical tone)の特徴がプロミネンスにどのような影響を与えるかという視点である。音調言語(tone language)や語ピッチ言語(Word-pitch language)では、語彙的トーンの特徴がイントネーション構造に強く影響する。この研究では、日本語を対象として、語彙アクセント(lexical accent)がプロミネンスの知覚にどのような影響を及ぼすかを検討する。

日本語において単語のアクセント型を判断する場合、2種類の情報を使うことができる。ひとつは音響情報で、有核型(accented)は無核型(unaccented)よりも F0 ピーク値が高く、急激な F0 下降を伴う。もう 1 つはレキシコンに記載された音韻的なアクセント情報である。通常、音節や語の知覚上のプロミネンスはその F0 変動幅(F0 excursion size)に依存し、それが大きければ大きいほど知覚されるプロミネンスも大きくなる(上に挙げた文献を参照)。これを前提にすれば、有核型の語は無核型の語よりも知覚上のプロミネンスは大きくなると予測される。

本研究では、この予測が当たらず、アクセント型の違いはプロミネンスを知覚する上では正規化されることを示す。Shinya (2005)では、自然な発話(音響的なアクセント情報と語彙音韻的なアクセント情報の両方が利用可能な発話)を用いて、アクセント型の正規化の存在を示した。Shinya (2006)では、音響情報からアクセント型が判別できないように F0 形状を曖昧に操作した発話も含めて実験を行った。これにより語彙音韻情報のみによりアクセント型の正規化が可能かどうかを検討し、これを支持する結果を得た。ここでは Shinya (2006)の実験を紹介する。

3. 実験

3.1 実験の手順

(1)に示した4つの文を録音し(発話者は筆者)、それぞれの第1アクセント句(accentual

phrase)(「稲森の」、「稲村の」)のF0形状を音響上の有核型と音響上の無核型の間で6段階で変化させた。第2アクセント句はF0形状を操作することはせず、F0ピーク値の高さのみを6段階に変化させた。第1アクセント句6刺激と第2アクセント句6刺激を組み合わせ、1文36個、計144語の刺激音を作成した。被験者はこれらの刺激音を聞き、第1単語と第2単語のうちどちらがより強調されて聞こえるかの判断を行った。得られたデータから、probit analysis(回帰分析の一種)によって第1単語と第2単語が同じ知覚上のプロミネンス(強調の度合い)を持つときの第1ピークと第2ピーク値を推定し、この値をraw dataとし分析した。以下に実験手順を刺激音の作成と実験の実施に分けて見ていく。

- (1) a. 有核-有核 (aa)
稲森の兄嫁がいない。
- b. 有核-無核 (au)
稲森のお土産が消えた。
- c. 無核-有核 (ua)
稲村の兄嫁がいない。
- d. 無核-無核 (uu)
稲村のお土産が消えた。

3.2 刺激音の作成 (PSOLA resynthesis)

Praat には PSOLA (Pitch Synchronous OverLap and Add)と呼ばれるアルゴリズムを用いた音声合成プログラムが装備されている。このプログラムを用いることにより、録音した発話の F0 形状を操作した再合成音を作ることができる。得られる音声の質も高い。具体的なやり方について、まずは一つ一つの操作をクリックしながら行う「手動」の方法を見た後、その操作をほぼ自動化する、スクリプトを使った方法を解説する。

まずは手動による刺激音作成であるが、以下のような手順を踏むことになる。この中で、F0 形状を stylize するのは、刺激音のコントロールを容易にすると同時に、F0 形状の特性を制限し、実験結果を解釈可能にするためである。

1. 再合成を行う音声を Read menu から Object window 上に読み込む。
2. To Manipulation... - OK で Manipulation ファイルが作られる。
3. Edit で編集画面を表示させ、Pitch menu - Stylize pitch で F0 形状を簡略化あるいはトーンのみによる表現に変える。各トーンは Praat 上では pitch point と呼ばれる。
4. pitch point を操作して、F0 形状を変化させる。変化した音声は下方のバーをクリックすることで聞くことができる。pitch point を正確に操作するには、Pitch menu - Remove pitch point(s) で当該の pitch point をいったん消去し、Add pitch point at... で時点と周波数を指定し再付与する。
5. Object window に戻り、Manipulation ファイルを選択し、Get resynthesis (PSOLA)。できた音声ファイルを Write menu から保存。

このように、刺激音の作成を手動で行うことは、個々の操作は難しくないにしても、多くの刺激音を作るとなると煩雑であるし、ミスが起こりやすくなる。このような時、Praatのスクリプト機能を使うことにより、手動で行う操作を自動化することができる。以下では2つのスクリプトを見る。これらのスクリプトを使うと、あらかじめstylizeしておいたManipulationファイルを基にして、複数の刺激

音を一気に作ることができる。

スクリプト1を以下に示す。内容については、スクリプト各行右の解説をご覧ください。このスクリプトは、あらかじめstylizeしておいたManipulation aaという名前のファイル(「稲森の兄嫁がない」)から、第2単語「兄嫁」のF0ピークの値を200Hzから10Hzずつ段階的に下げて6つの刺激音を作り、指定したディレクトリに保存する。

```
1  outdir$ = "output\"           作成された音声ファイルを格納するディレクトリ
2  for x from 0 to 5             Endfor(15行目)までを6回繰り返す(ループ)
3      select Manipulation aa    Manipulation aaを選択
4      Extract pitch tier        pitch tierを別個に作成
5      Remove point... 5        5番目のpitch pointを削除
6      f0 = 10*x                 1段階でのF0の変化幅を変数f0として定義
7      Add point... 0.7931 200-f0 pitch pointを0.7931秒の地点に200-f0 Hzで追加
8      select Manipulation aa    Manipulation aaを選択
9      plus PitchTier untitled   作られたpitch tier(4行目)を追加で選択
10     Replace pitch tier         pitch tierでされた操作結果をManipulation aaに置き換え
11     select Manipulation aa    Manipulation aaを選択
12     Get resynthesis (PSOLA)   再合成
13     Write to WAV file... 'outdir$'aa'x'.wav WAVEファイルとして保存
14     select PitchTier untitled pitch tierを選択
15     Remove                    pitch tierを削除
16 endfor
17 Clearinfo
18 print "Your stimuli have been created!!" } 刺激音作成終了のコメント
```

図1 スクリプト1

手動で行う場合と違い、スクリプトではManipulationファイル内のpitch tierの情報を独立したpitch tierとしていったん取り出す必要がある。取り出したpitch tier上でF0の操作を行った後、操作結果をManipulationファイルに貼り付けている。スクリプト1を実行した結果作られた刺激音のF0形状を図3aに示す。

スクリプト2は、3つのpitch pointsそれぞれを段階的に変えるより複雑なスクリプトである。3つのpitch pointのうち、最初の2つは第1単語のF0ピーク(有核型ではH*, 無核型ではH)とそれに続くトーン(有核型では+L, 無核型ではspread H, Sugahara 2003)に相当する。3つ目のpitch pointは第2単語のF0ピーク(H*)である。このスクリプトを実行してできる刺激音は、第1単語に関しては有核型から無核型の6段階の連続体であり、第2単語に関しては6段階のF0ピーク値の連続体である。これは実際の実験で刺激音を作るために用いたスクリプトである。このスクリプトについてここで詳述する余裕はないが、基本的にはスクリプト1を拡張したものであり、内容的にそれほど飛躍があるわけではない。このスクリプトは大きく3つの部分に分けられる。3-13行目、14-24行目、26-38行目がそれぞれ第1、第2、第3pitch pointの処理に当たる。第3pitch pointのループは第1、第2pitch pointのためのループの中に埋め込まれており、これにより第1、第2pitch pointのそれぞれの組み合わせに対して6つの第3pitch pointの値が与えられる。このスクリプトを実行した結果得られた刺激音のF0形状は図3bに示した。

4. 実験の実施 (Experiment MFC)

刺激音ができたら、次はその刺激音を使って実際に実験をするわけであるが、その際にもPraatを使うことができる。PraatにはExperiment MFC (Multiple Forced Choice)という実験実施用のプログラムが装備されている。以下ではその使い方を解説する。

```

1  outdir$ = "output\"
2  for x from 0 to 5          第1, 第2pitch pointのループ
3      select Manipulation aa
4      Extract pitch tier
5      Remove point... 2
6      align_point2 = 0.0026667*x  1段階でのalignmentの変化幅を変数align_point2として定義
7      f0_point2 = 2.833333*x      1段階でのf0の変化幅を変数f0_point2として定義
8      Add point... 0.213-align_point2 196-f0_point2
9      select Manipulation aa
10     plus PitchTier untitled
11     Replace pitch tier
12     select PitchTier untitled
13     Remove
14     select Manipulation aa
15     Extract pitch tier
16     Remove point... 3
17     align_point3 = 0.0085*x
18     f0_point3 = 4.83333*x
19     Add point... 0.428+align_point3 122+f0_point3
20     select Manipulation aa
21     plus PitchTier untitled
22     Replace pitch tier
23     select PitchTier untitled
24     Remove
25     for y from 0 to 5      第3pitch pointのループ
26         select Manipulation aa
27         Extract pitch tier
28         Remove point... 5
29         f0_point5 = 14*y
30         Add point... 0.7931 120+f0_point5
31         select Manipulation aa
32         plus PitchTier untitled
33         Replace pitch tier
34         select Manipulation aa
35         Get resynthesis (PSOLA)
36         Write to WAV file... 'outdir$'aa'x'y'.wav
37         select PitchTier untitled
38         Remove
39     endfor
40 endfor
41 clearinfo
42 print "Your stimuli have been created!!"

```

図2 スクリプト2

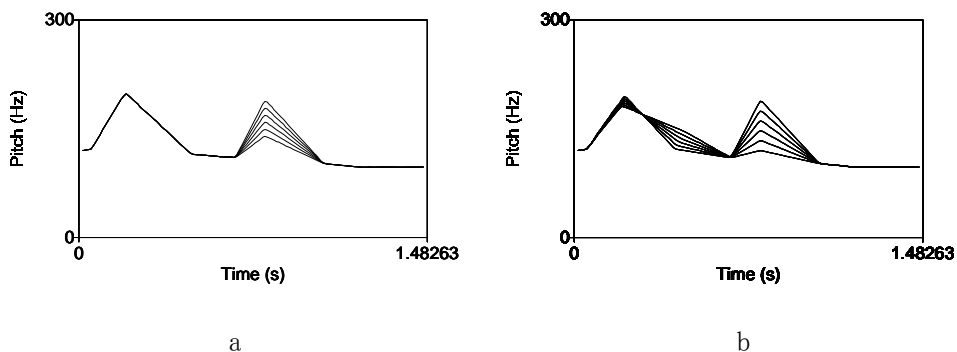


図3 スクリプト1, 2を実行して作成された刺激音のF0形状

```

1 "ooTextFile"
2 "ExperimentMFC 2"
3 "stimuli/"
4 ".wav"
5 carrierBefore = ""
6 carrierAfter = ""
7 initialSilenceDuration = 1.5 seconds
8 interStimulusInterval = 0
9 numberOfDifferentStimuli = 36
10 "aa11"
11 "aa12"
12 "aa13"
13 .....
14 "aa64"
15 "aa65"
16 "aa66"
17 numberOfReplicationsPerStimulus = 5
18 breakAfterEvery = 72
19 randomize = <PermuteAll>
20 startText = "Click to start."
21 runText = "Which of the two words is
22 given more importance by the
23 speaker?"
24 pauseText = "End of block of
25 trials. Have a short break. Click to
26 proceed."
27 endText = "The practice session is
28 over."
29 Tell experimenter you're done."
30 numberOfResponseCategories = 2
31 0.1 0.4 0.35 0.65 "1" "1"
32 0.6 0.9 0.35 0.65 "2" "2"
33 numberOfGoodnessCategories = 0
34 0.25 0.35 0.10 0.20 "1 (poor)"
35 0.35 0.45 0.10 0.20 "2"
36 0.45 0.55 0.10 0.20 "3"
37 0.55 0.65 0.10 0.20 "4"
38 0.65 0.75 0.10 0.20 "5 (good)"

```

図4 Experiment MFCのパラメータ設定

単語の F0ピーク(P2)の平均値を縦軸に示されている。横軸の値は第1単語の第1トーン(H*あるいはH)の値を示しており、中間部がアクセント型に関して音響的に曖昧な刺激音である。グラフ上の直線は $y=x$ の関数である。もし P1 と同等のプロミネンスを持つ P2 の値が P1 のそれと同じである場合には、P2 の値はこの関数上に位置するはずである。また、各平均値とこの関数との間の距離はそのまま P1 と P2 のピーク差を表している。

図5のグラフからは次の3つのパターンを読み取ることができる。まず第1に、すべての平均値がこの関数よりも低い値を示している。これは連続する2つの F0 ピークが同じプロミネンスを持つ場合、P2 は P1 よりも物理的に低い F0 値でなければならないことを意味し、Pierrehumbert (1979)の言う自然下降の正規化が働いているものと思われる。第2に、第1単語が一定の場合、第2単語が有核型の時(黒の三角と四角)には、それが無核型の時(白の三角と四角)に比べて P2 の平均値が高くなり、P1 と P2 の差が小さくなっている(統計的にも有意 $F(1,23)=71.383, p<0.001$)。これは聞き手が、有核型の語が無核型の語よりも F0 変動幅が大きいことを知識として持っていて、そのプロミネンスを実際よりも低く見積ると解釈される。つまり、語彙アクセントを知覚上正規化していると考えられる。第3に、第2単語が一定の場合、第1単語が有核型の時(白

MFC Experimentでは、テキストファイル上で必要なパラメータを設定し、そのファイルをPraat上でRead menuから読み込むことで実験を実施できる。図4がテキストファイルの1例である。やり方は単純である。紙面の都合上すべてのパラメータについてここで取り上げられないので、いくつか選んで述べることにする。3-4行目で刺激音が含まれるディレクトリとファイル形式を指定する。10-16行目には刺激音のファイル名を入力する。19行目は刺激提示のrandomizationの方法を指定する部分で、いくつかのパターンが指定できる。〈PermuteAll〉はすべての刺激音をrandomizeする最も単純な方法である。21-29行目では、実験の初め、short break、終わりで表示するテキストを指定する。30-32行目では、回答用のボタンの位置と大きさを指定する。Praatでは1つの刺激に対して2種類の回答を記録することが可能であり、33-38行目では2種類目の回答のボタンの指定を行う。本実験ではこの部分は使用していない。

実験が終わったら、データはObject window上のExtract results - Write to text fileで保存できる。

5. 実験結果と考察

Probit analysisの結果を図5に示した。第1単語の有核-無核の連続体が横軸(P(eak)1)、そのP1と同等の知覚上のプロミネンスをもつ第2

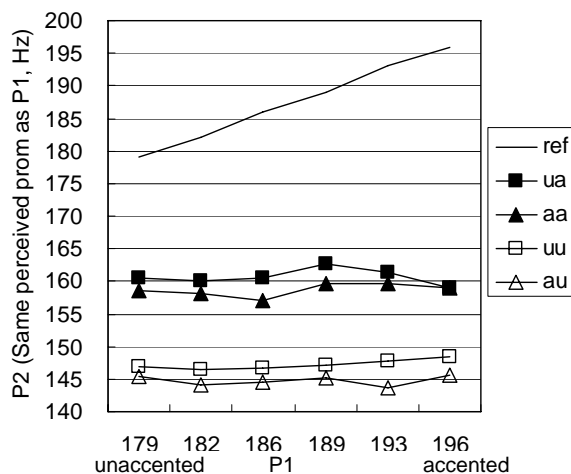


図5 P1と同等の知覚上のプロミネンスを持つP2の平均値

は全く同じであるということである。これは、聞き手は音響的なアクセント情報が存在しない場合でも、レキシコンに記載された語彙アクセントの音韻情報を使って、プロミネンス知覚の正規化を行っていることを示唆する。また、第2のパターンは第3のパターンよりも正規化の効果が大きい。これは第2単語は音韻情報とともに音響情報も利用可能であったからと考えられる。

これらの実験結果が示唆する1つの理論的含意は、日本語においてアクセントのあるモーラを音韻的に強い位置(phonologically strong position)とは言えないのではないかということである。少なくともそう主張するための心理的実在性はこの実験では見つからなかった。

6. まとめ

Praatを使った日本語イントネーションのプロミネンス知覚に関する研究について概説した。Praatを使うことにより、これまで高価な設備が必要だった合成音声を使った研究が「無料で」行うことができるようになったことは大きい。これにより、音声・音韻研究の益々の発展を期待したい。

7. 引用文献

- Boersma, P. and Weenink, D. (1992-2005) *Praat: Doing phonetics by computer*. [Computer program].
- Gussenhoven C. and Rietveld, A. C. M. (1988) "Fundamental frequency declination in Dutch: testing three hypotheses," *J. Phonetics* 16, 335-369.
- Gussenhoven, C., Repp, B. H., Rietveld, A. C. M., Rump, H. H. and Terken, J. (1997) "The perceptual prominence of fundamental frequency peaks," *JASA*, 102, 3009-3021.
- Pierrehumbert, J. (1979) "The perception of fundamental frequency declination," *JASA* 66, 363-369.
- Shinya, T. (2005) "Lexical accent effects on the perception of fundamental frequency peaks in Japanese." The 150th Meeting of the Acoustical Society of America, Minneapolis, USA.
- Shinya, T. (2006) "Lexical accent status affects perceived prominence of intonational peaks in Japanese," *Proceedings of Speech Prosody 2006*. Dresden, Germany, 89-92.
- Shinya, T. (forthcoming) *The Role of Lexical Contrast in the Perception of Intonational Prominence in Japanese*. Doctoral Dissertation, University of Massachusetts, Amherst.
- Sugahara, M. (2003) *Downtrends and Post-FOCUS Intonation in Tokyo Japanese*. Doctoral Dissertation, University of Massachusetts, Amherst.
- Terken, J. (1991) "Fundamental frequency and perceived prominence of accented syllables," *JASA* 89, 1768-1776.

と黒の三角)には、それが無核型の時(白と黒の四角)に比べてP2の値が低くなり、P1とP2の差が大きくなっている($F(1,23)=7.625, p=0.013$)。これは第2のパターンと本質的に同じ現象と見ることが出来る。つまり、一定の第2単語のプロミネンスに対して、第1単語がそれと同等のプロミネンスを持つ場合、有核型の語のプロミネンスは実際のF0特性よりも低く判断されるので、無核の語と同等のプロミネンスを持つためにはより高いF0ピーク値が必要になる。この結果、P1とP2の差が大きくなるのである。

第3のパターンで重要なのは、uaとaa(黒の三角と四角)、そしてuuとau(白の三角と四角)は、それぞれF0形状として