

母語話者日本語音声を対象にした各種文脈における 語頭 2 モーラの F_0 上昇に関する分析

吉澤 風希・峯松 信明・齋藤 大輔（東京大学大学院）
{yoshizawa, mine, dsk_saito}@gavo.t.u-tokyo.ac.jp

1 はじめに

日本語東京方言における 1 型以外の単語の孤立単語発声において、「第 1 モーラから第 2 モーラにかけてピッチが上昇する」と説明されることが多い。音声において心理量であるピッチに対応する物理量は基本周波数 (F_0) であるが、第 2 モーラが長母音の二拍目または撥音である場合には F_0 の上昇が抑えられる。これは異音的変動と呼ばれる (Hattori, S., 1954; Lei, 2016; Pierrehumbert, J. B., Beckman, M. E., 1988)。

近年、日本語学習者用に任意の入力文を東京方言で読上げる際に必要なピッチ制御を表示する、Online Japanese Accent Dictionary (OJAD) と呼ばれる web 教材が運用されている (峯松ほか, 2013)。その機能の一つに韻律読み上げチュータ「スズキクン」がある。これは任意の文章に対し東京方言での発声に必要な韻律制御を可視化し、その通りに読み上げた手本となる合成音声を提供する機能である。スズキクン利用者の利用による発音の向上は、実験的に確認されている (峯松・平野・中村, 2016)。また弁論大会参加者の練習にも利用されるツールになっている。OJAD では、ピッチ制御をシンボルを用いて離散的に表示するのではなく、入力テキストに対する自動解析結果のみを利用し、仮想的なピッチパターンに対応する曲線を用いて示している。OJAD におけるピッチパターン表示は異音的変動を考慮せず、語頭の 2 モーラで常にピッチが明確に上昇するカーブを描く。本研究ではこのピッチパターン表示をより実際の現象に近づけることを念頭に置き、任意の文入力における異音的変動に対応できるよう、各種文脈における語頭 2 モーラにおける F_0 上昇について音響分析を行なった。

2 分析実験

第 2 モーラが長母音の二拍目または撥音である単語の、語頭での F_0 上昇の変化を分析するために実験を行った。実験は、韻律的なコンテキストを考慮するために、該当単語の直前が無音、L トーン、H トーンの 3 つの場合で行った。

注目する単語が含まれる文と、比較対象となる単語が含まれる文をそれぞれ読み上げた音声を取録し、音声の F_0 のモーラごとの代表値を求め、語頭でのアクセント上昇を F_0 の変化として定量的に分析した。更に注目単語の直前が無音の場合だけでなく、L である場合と H である場合についても音声を録音し、同様の評価を行った。これは (Lei, 2016) では、単独での発声のみに言及していたが、それに加えて、文中に置かれた時の発声での変化を観察するためである。このように注目単語の直前として、無音・L・H の 3 パターンを想定し、これらを以下ではコンテキストと呼ぶ。

表 1: 実験文リスト。1 モーラ目の母音が異なる 2 文のペアを用意した。

	a	i	u	e	o
2モーラ目が通常拍	ラムネを見つける	千鳥ヶ淵を見学する	熊本に向かう	セミプロになりました	駒沢大学を見学する
2モーラ目が長音	ラー油を見つける	地域おこしを見学する	空港に向かう	成人になりました	工業大学を見学する
	a	i	u	e	o
2モーラ目が通常拍	窓枠を見つける	締め出しをする	ブルジョアを訪ねる	ゲレンデが見える	小麦粉を破棄する
2モーラ目が撥音	万力を見つける	申告をする	文豪を訪ねる	限界が見える	婚約を破棄する

2.1 実験文

第2モーラが長母音の二拍目である単語と撥音である単語それぞれについて、実験文を以下のように構成した。まず、上記の条件を満たす単語から第1モーラの母音がそれぞれア・イ・ウ・エ・オになるように計5単語を選出した。さらに比較対象の単語として、これらの単語と同じモーラから始まり、単語長とアクセント型が同じである単語の内、第2モーラが有声音となる単語を比較対象として選出した。このように5ペアの単語を選びだし、各ペアの単語を文頭に持ち後続の部分は同一である文のペアを5ペア構成した。これらの5ペアの文に「無事 (LH)」と「今朝 (HL)」を文頭に置いた場合を加え、3種のコンテキストを反映した文を実験文とした。実験文は総計 $2 \times 5 \times 3 = 30$ ペアとなる。表1に文頭に何も置かない場合の実験文を示す。

2.2 実験データ

東京方言話者である20代男性5名が実験文を読み上げた音声を実験データとした。読み上げの際には、なるべく自身にとって読みやすいスピードで、かつ文ごとにスピードを変えないように指示した。被験者とは事前に個々の単語について正しいアクセントを確認したが、例となる音声は聴かせなかった。また、「無事」や「今朝」が文頭に付く場合に、ポーズを置かないように指示をした。

録音はサンプリング周波数48kHz、量子化bit数が16bit、モノラルで行い、これを16kHzにダウンサンプリングしたものを利用した。データ数は、60文を5人が読み上げたため300発声である。

2.3 分析手法

ピッチの上昇量を定量的に比較する為に、実音声から $\log F_0$ の上昇量を求めた。 F_0 抽出には SPTK (SPTK, 2017) を利用した。SPTK での F_0 抽出は、SWIPE アルゴリズムを用い、フレームシフト 80、SWIPE 閾値 0.3、最小 F_0 を 50 Hz、最大 F_0 を 400 Hz として行った。

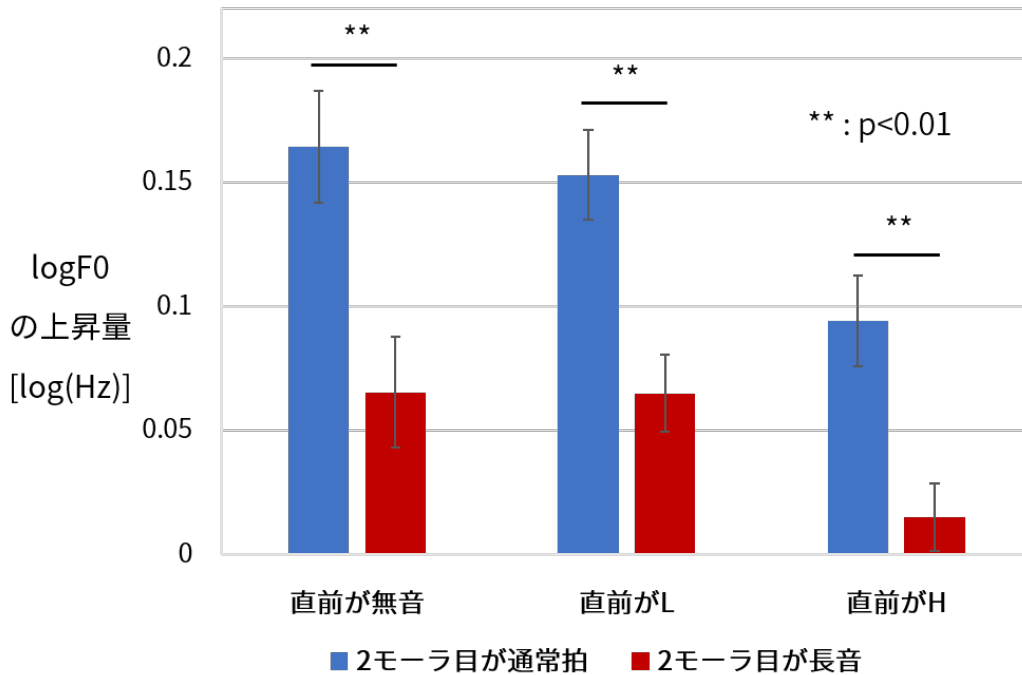


図 1: 長母音の有無と F_0 上昇量の関係 (エラーバーは標準誤差を表す)。

また、Julius (Lee & Kawahara, 2009) の音素セグメンテーションキットによって音声と音素の強制アラインメントを取り、実験文を構成する音素に対応する時間情報を得た。時間情報を元に、文頭の単語、あるいは「無事」、「今朝」の後にある注目している単語について、第1モーラの母音部の開始時刻と第2モーラの母音部の終了時刻を求め、これら2時刻での $\log F_0$ の差分を計算した。 F_0 の検出に失敗した部分がある場合は、 F_0 が存在している時刻をモーラ内で探索し、それぞれ開始時と終了時に最も近い時点での値を採用した。

分析を行う際には、注目単語の第1モーラの母音で区別せずにデータを利用した。

3 実験結果

3.1 第2モーラが長母音の二拍目である単語

第2モーラが長母音の二拍目である単語についての分析結果を図1に示す。

平均値を見ると、直前が無音・L・Hのいずれのコンテキストであっても、第2モーラが長母音の単語はそうでない単語より $\log F_0$ の上昇量が有意に小さかった ($p < 0.01$)。

コンテキストが無音やLである場合、第2モーラが長母音の二拍目である単語の語頭での F_0 上昇は、抑えられるものの完全に消失してはいなかった。一方コンテキストがHであると、 F_0 の上昇はほぼ消失していた。コンテキストごとに見た場合、傾向としての差は存在したが5%の有意差は示されなかった。

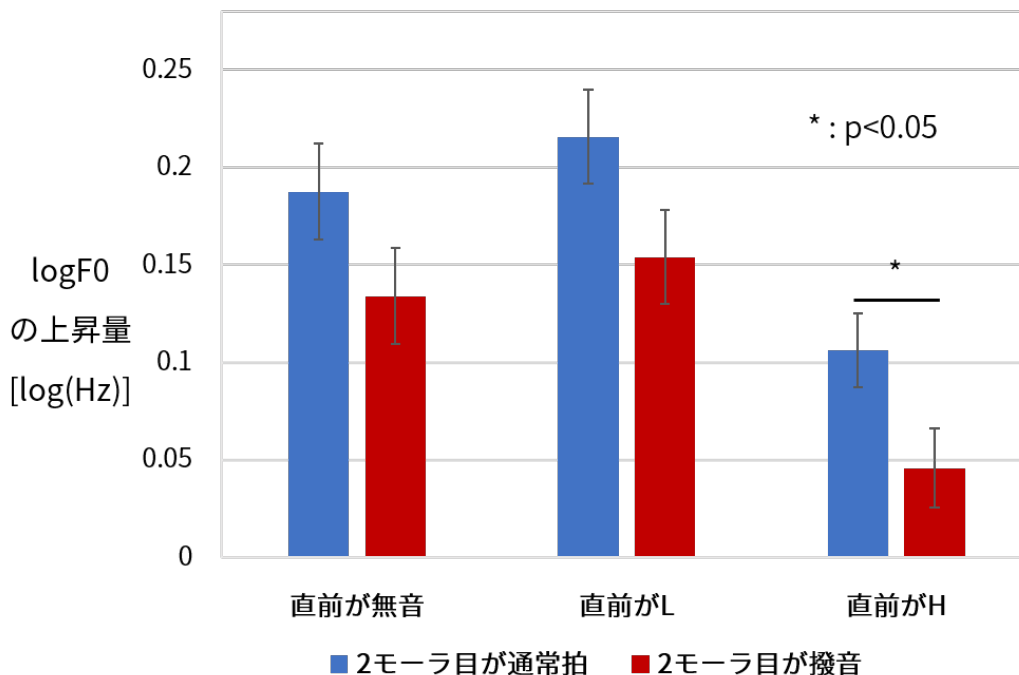


図 2: 撥音の有無と F_0 上昇量の関係 (エラーバーは標準誤差を表す)。

3.2 第2モーラが撥音である単語

第2モーラが撥音である単語についての分析結果を図2に示す。

平均値を見ると、 F_0 の上昇は抑えられている傾向にある。しかし分散が大きいため、直前が無音・Lのコンテキストである場合には、5%の有意差が示されなかった。一方、直前がHコンテキストである場合には、5%の有意差が示された。

3.3 母音ごとの比較

第2モーラ目が撥音となっている単語について、 F_0 パターンを観察した。その結果、第1モーラの母音によって語頭のアクセント上昇の抑えられ方が変化する傾向にあることを見いだせた。第1モーラの母音ごとについての分析結果を図3に示す。

第1モーラの母音が「イ」である場合は、コンテキストに依らず $\log F_0$ の上昇が抑えられる傾向が見られる。また、第1モーラの母音が「オ」である場合は、単語の直前が無音かHコンテキストであると、上がり方が抑えられる傾向にある。

なお、いずれの場合も5%の有意差は示されていない。

4 考察

第2モーラが長母音の2拍目または撥音である単語について、全体として F_0 の上昇は抑えられる傾向にあった。

またコンテキストに注目すると、単語直前がHであると、語頭での F_0 上昇は他のコンテ

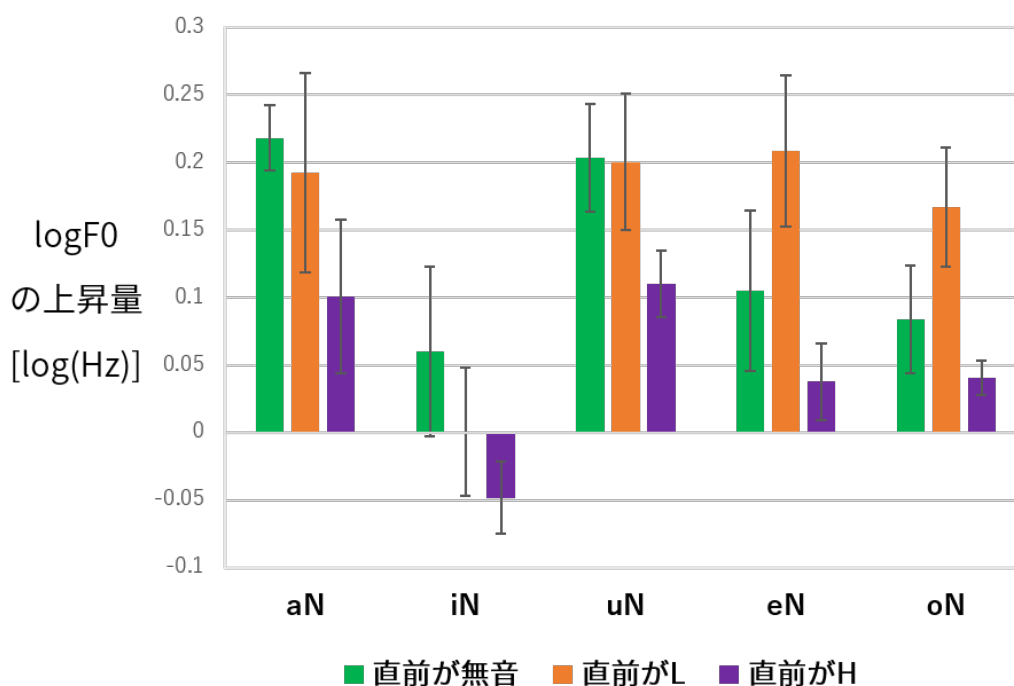


図 3: 撥音の直前にある母音と F_0 上昇量の関係 (エラーバーは標準誤差を表す)。

クスト下と比較して、大きく抑えられている。第 2 モーラが長母音の 2 拍目または撥音である単語の直前に、「無事 (LH)」のような H で終わる語が存在していると、当該単語は 1 モーラ目から 2 モーラ目へかけてのピッチの上がりがないように読まれる。つまり単純にアクセントを並べた場合に LH **L** HHH となる文が、LH **H** HHH と読まれていることを意味している。この傾向は第 2 モーラが長母音の 2 拍目である単語では特に顕著で、語頭での F_0 上昇はほぼ消失していた。

また、単語の 2 モーラ目が撥音であることの語頭でのアクセント上昇への影響は、1 モーラ目の母音によって差が生じる可能性が示唆された。第 1 モーラの母音で区別せずに分析を行った場合には有意差は示されなかったが、データを十分な数用意して第 1 モーラの母音ごとに分析を行えば、第 1 モーラの母音によっては明確な有意差が示される可能性がある。

本実験では、 F_0 上昇量の分散が大きくなってしまっていた。その原因の 1 つにアクセントの混在が考えられる。被験者には「無事ラムネを見つける。」といった「無事 (LH)」から始まる文を読み上げてもらった。しかし、被験者に何も指示をせずにこれらの文を発声させると、LH **H** HHH…と読む被験者と LH **L** HHH…と読む被験者が混在していた。「ラムネを」という部分のを単独で読み上げると、そのアクセントは LHHH である。しかしこれを「無事」という語の後に続けて読んだ場合に、「ラムネを」の語頭が L になるかどうかは、話速などの各個人の発話スタイルに依存する。そのため、アクセントに差異が生じる。

これらの差異を区別していないことが、ノイズになっている可能性がある。ただ、これらの差異をコントロールするためにどちらかのアクセントで読むように指示をすると、発声の自然性が失われる可能性が生じる。そのため、特に被験者には指示を出さずに読んでもらい、後でどちらのアクセントで読まれているかをラベリングし、データを分けて分析をすべきだと考えられる。

5 おわりに

本研究では、各種文脈下における語頭2モーラでの F_0 上昇に関する音響分析を行った。分析実験の結果、第2モーラが長母音の二拍目または撥音である場合には語頭での F_0 の上昇が抑えられる、異音的変動と呼ばれる現象が実際に起きていることが確認できた。また、この異音的変動は、単語直前のアクセントの影響を受けていることを実験的に検証した。この帰結より、OJAD のスズキクンにおけるピッチパターンをより現実のものに近づけるためには、異音的変動を考慮して、特定の文脈・単語において語頭2モーラでのピッチの上昇を緩やかにして提示するべきであると考えられる。

参考文献

- Hattori, S (1954) “The Japanese accent in a view of phonology.” *Kokugo kenkyu* 2(2), 37-65.
- Lee, A., & Kawahara, T. (2009) “Recent development of open-source speech recognition engine julius.” In *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference* (pp. 131–137).
- Lei, H.-Y. (2016) “An acoustic analysis on the allotonic variation of the initial rise in Tokyo Japanese in native speakers and learners.” In *Proc. International Symposium on Applied Phonetics*.
- Pierrehumbert, J. B., Beckman, M. E. (1988) *Japanese Tone Structure*. The MIT Press.
- Speech Signal Processing Toolkit (SPTK)*. (2017). <http://sp-tk.sourceforge.net/>. (最終アクセス日 2018 年 1 月 27 日).
- 峯松 信明・中村新芽・鈴木雅之・平野宏子・中川千恵子・中村則子・田川恭識・広瀬啓吉・橋本浩弥 (2013) 「日本語アクセント・イントネーションの教育・学習を支援するオンラインインフラストラクチャの構築とその評価」 『電子情報通信学会論文誌 D』 96: 10, 2496–2508.
- 峯松 信明・平野 宏子・中村 則子 (2016) 「OJAD を用いた日本語韻律トレーニングによる自然性向上に関する実験的検証」 『電子情報通信学会技術研究報告音声 (SP)』 115: 392, 13–18.