2017(平成 29)年度

第 31 回

日本音声学会全国大会予稿集

Proceedings of the 31st General Meeting of the PSJ

2017 年 9 月 30 日・10 月 1 日

September 30 - October 1, 2017

東京大学本郷キャンパス 工学部2号館

Engineering Building 2, Hongo Campus

The University of Tokyo

日本音声学会

The Phonetic Society of Japan

第 31 回日本音声学会全国大会を迎えて

第 31 回日本音声学会全国大会を東京大学で開催するにあたり，様々な便宜をお取り計らい下さった大久保達也東京大学大学院工学系研究科長，大会運営委員長の労を快くお引き受け頂いた峯松信明東京大学教授に，心から感謝申し上げます.

創立 91 周年を迎えて日本音声学会は，会員諸氏の活発な研究活動を基盤として，教育や医療，文化，ビジネス分野などで要求される実践的「音声学」の育成にも積極的に貢献していきたいと考えております.

第 31 回全国大会ではその一環として二つのワークショップ,「プロソディ研究のための方法論：コーパス・生理・文タイプ」と「英語教員を目指す人への英語音声指導法」を企画いたしました．これを契機に研究と実践をつなぐ議論をさらに深めていきたいと期待しております.

これに関連して，日本音声学会では，言語教育に携わる教員養成課程における音声学の必須化を要望する提言「指導要領に定める英語音声教育実現のための提言」を文部科学大臣に提出いたしました．提言の文言は学会ホームページに掲載いたしました

(http://www.psj.gr.jp/jpn/wp-content/uploads/2017/06/20170525.pdf).

第 31 回全国大会では，口頭発表４会場，ポスター発表１会場に分かれて，多様なテーマについて活発な討論と情報交換が二日間にわたって行われます．全国大会の締めとなる公開講演では，柏野牧夫博士（日本電信電話株式会社コミュニケーション科学基礎研究所）から「錯聴から音声知覚のメカニズムを探る」，出口利定博士（東京学芸大学学長）から「乳幼児における聴知覚の発達」と題したご講演を頂きます．音声の知覚機構をめぐる謎，乳幼児が生後瞬く間に話し言葉を理解できるようになる発達の謎を分かり易く解きほぐして下さると楽しみにしております.

今大会から全国大会予稿集を USB や冊子媒体で配布する方法を改め，会員マイページを介して事前に電子公開することにいたしました．電子公開は情報発信の範囲拡大や迅速化・利便性の向上，経費削減など多くの利点があることをぜひご理解いただき，戸惑いもあろうかと思いますが，ぜひご協力をお願い致します．また，会誌『音声研究』の J-STAGE を介した公開も８月中に完成しました.『音声研究』の冊子体発行を継続しつつ，電子ジャーナル化に向けた検討も慎重に進めて参りたいと考えております.

最後に，活発な研究発表と熱心な意見交換に参加して下さる皆々様，大会の企画・運営を支えてくださる皆様に深謝し，ご挨拶とさせて頂きます.

2017（平成 29）年 9 月

日本音声学会会長 今泉 敏

公開講演1（PL1）

錯聴から音声知覚のメカニズムを探る

柏野　牧夫（NTT コミュニケーション科学基礎研究所，東京工業大学）
kashino.makio@lab.ntt.co.jp

1. はじめに

一般に，音の知覚内容は，耳に入ってくる音の物理特性とは多かれ少なかれ乖離している．この乖離の顕著なものが錯聴（auditory illusion）と呼ばれる（Bregman, 1990; 柏野, 2010; 柏野, 2012; Warren, 2008; ウェブサイト「イリュージョンフォーラム」http://www.kecl.ntt.co.jp/IllusionForum/index.html）．錯聴は単なる情報処理のエラーではなく，日常の環境で適切に知覚する上で有用な適応的意義を持っていることが多い．錯聴の特性を系統的に分析することによって，聴覚情報処理の原理を推測することができる．また，錯聴における音の物理特性と知覚特性の乖離は，脳活動の計測とうまく組み合わせれば，知覚が脳内のどこでどのように形成されるのかを探索する上で有効な武器になる．

本稿では，特に音声知覚（speech perception）に関わる錯聴を 2 タイプ取り上げ，そこから見えてくる脳内情報処理のメカニズムについて論じる．1 つめは，音響信号の劣化にもかかわらず元の発話内容が知覚される「知覚的修復（perceptual restoration）」，2 つめは，同一の音パターンが反復呈示されると知覚内容が次々と変化する「多義的知覚（multistable perception）」である．このうち，錯聴を利用して音声知覚の脳内メカニズムに迫る研究の一例として，多義的知覚に関わる脳活動を分析した我々の研究を紹介する．そして，音声知覚の情報処理原理について，並列・階層的な予測符号化という観点から考察する．

2. 知覚的修復

音声知覚は音響信号の劣化に対して驚くほど頑健である．信号の劣化があたかも「修復」されたように知覚され，かなりの程度まで元の発話内容が聞き取れる．このような知覚的修復（perceptual restoration）には様々なものがあるが，その代表例が音素修復（phonemic restoration）である（Warren, 1970）．まず，文章を読み上げたものを録音し，その頭から 100〜200 ms 程度の間隔ごとに音声を削除して無音にする（図 1a）．こうすると，発話内容が非常に聞き取りにくくなる．次に，音声を削除した部分に雑音を挿入する（図 1b）．雑音の周波数帯域は音声よりも広く，音圧レベルは音声よりも高くする．こうすると，雑音の背後で，削除されたはずの音声が修復され，滑らかにつながって聞こえる．雑音があろうがなかろうが，同じ時間長の音声が削除されたことに変わりはないが，聞こえ方は劇的に違う．雑音が挿入されると，単に滑らかに聞こえるだけでなく，発話内容が聞き取りやすくなる．音素修復の効果は劇的で，最適な条件下では，どの音素が欠落しているのかわからないほどである．また，「本当は音声が削除されている」という事実を知っていても，この錯聴を阻止することはできない．意識的に欠落部分を推測するのではなく，自動的に「聞こえてしまう」のである．

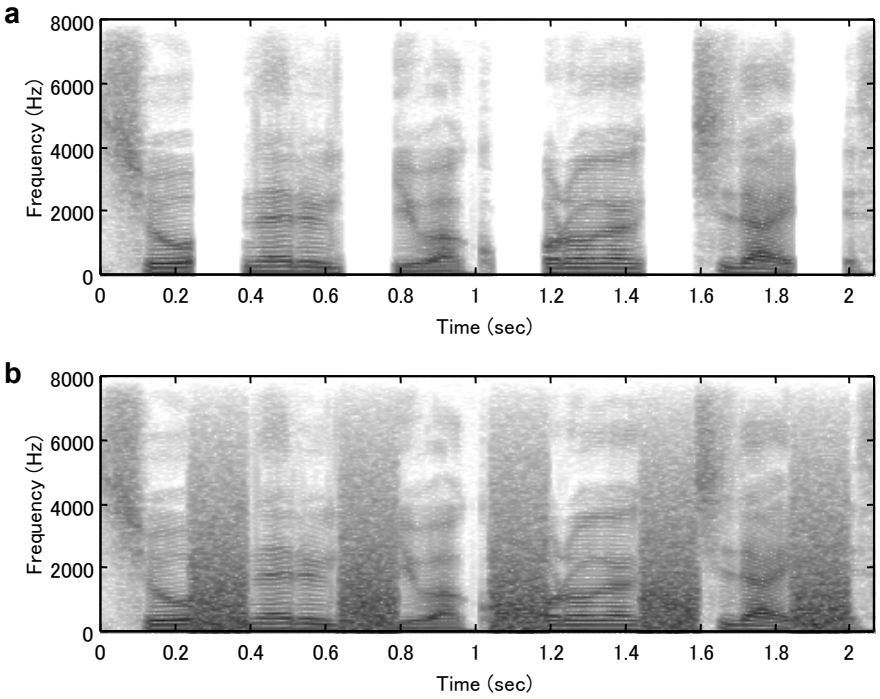


図 1: 音素修復. a: 省エネルギーは心がけ次第です」という言語音声を一定時間ごとに削除(無音置換)したもの. 下段:上段の無音置換の部分に雑音を挿入したもの. 縦軸は周波数, 横軸は時間, 濃いところはエネルギーが大きい部分を表す.

音素修復が生じるための音響的条件は，ひとことで言えば，「音声を削除した部分に挿入する雑音が，本来の音声がそこにあったとしても十分マスキングする（検知できなくする，隠す）ことができる特性を備えている」ということである．これを「マスキング可能性の法則」という．上記の例で，「雑音の周波数帯域は音声よりも広く，音圧レベルは音声よりも高くする」と述べたのは，雑音が音声をマスキングできるような特性を持たせるためである．別の言い方をすれば，音声が削除されている事実が検知されなければよいのである．したがって，音声と雑音の間に検知できるような時間的ギャップがある場合には，音素修復は妨げられる．

雑音で隠された方がよく聞こえるというのは，一見逆説的に思われるかもしれない．しかしこれは実は合理的である．ある人の話を聞いているときに，別の人が大きな咳をしたり，ドアをバタンと閉めたりして，音声の一部がマスキングされることは日常よくある．音素修復は，このように妨害音の多い環境で安定して音声を聞き取る上で大いに役立っている．脳は，マスキングされた部分に本来どのような音声があったかについての予測（仮説）を，その前後の情報から自動的に生成する．しかし，その部分にマスキングできるような雑音がなければ，「そこには音声がない」という明白な証拠になるので，「こういう音声があったはず」という予測は棄却される．したがって音素修復は生じない．そもそも修復が必要なのは，本来存在している音声が別の音でマスキングされた場合だけであって，

もともと切れている音声を補完する必要はない．この点，「隠された可能性があるときだけ補完する」というマスキング可能性の法則はきわめて理に適っている．

もう一点，音素修復の重要な示唆は，音素の知覚のための音響的特徴が，ある程度広い時間範囲に分散しているということである．信号削除の時間長や，その前後の音響信号を系統的に操作した実験によると，連続音声では各音素に関する情報が 200〜300 ms 程度の時間範囲に，隣接音素の情報が重畳する形で分散しており，それらの情報を統合することによって音素修復が実現されている（柏野, 1992）．

3. 多義的知覚

音の物理特性と知覚との乖離を如実に示す現象として，物理的には同一の音の聞こえ方が状況によって変化するというカテゴリーのものがある．これを一般的に多義的知覚（multistable perception）という．複数の音声を混合したものを聞くとき，注意の向け方によって異なった発話内容が知覚されるという選択的聴取も広い意味ではこのカテゴリーに分類できる．ここでは，同一の単語を反復すると知覚内容がどんどん変わっていくという現象を紹介する．まず，短い単語を録音する．例えば「バナナ」とやや早口で発声する．次に，それを切れ目なく反復再生する．すると，「バナナ」のはずが，「ナッパ」になったり，「ハナ」になったり，さらには 2 人の声に分かれたり，機械的な音が聞こえてきたりといった具合に，人によって中身は様々であるが，多くの場合，数分間聞けば知覚の変化が体験できる．変化の回数には個人差がある上に，知覚される内容もバラエティに富む人もいれば，比較的少数の聞こえ方が交互に現れるという人もいる．この現象自体は古くから知られていて，反復単語変形効果（verbal transformation; VT）と呼ばれている（Warren & Gregory, 1958）．

視覚でも，同一のパターンを長時間見続けているといくつかの見え方が切り替わる多義的知覚の現象が各種知られている．VT は，それらと共通点も相違点もある。「バナナ」を反復すると，「ナバナ」，「ナナバ」，さらには「バナ」「バナナバ」など音節のまとまり方に多義性ができる．このような多義性の中からある時点で一つを選んで知覚するという意味では視覚の例と共通であるが，視覚の例が空間的なパターンの解釈に関わる（動きを伴うパターンであっても時間的構造を持たない）ものであるのに対し，VT の方は時間的なパターンに関わる選択が本質的である．実際には，VT で知覚される内容は単に元の発声に含まれる音素の組み替えだけでなく，もともと含まれていない音素（「ハナ」とか「ナンパ」など）が聞こえることも珍しくない．音声を聞き取る際には，連続して発声された音声を音素，音節，単語などのまとまりに区切る時間的分節のプロセスと，分節されたパターンを脳内に蓄えられた言語的パターンと照合するプロセスとが不可欠であるが，VT は，これらのプロセスが特殊な人工的音声に対して作動したものと考えることができる．日常環境でも，発声が不明瞭であったり別の音で妨害されたりすれば，音声に含まれる情報が不十分で唯一の解釈に絞りきれないことがある．そういう場合，複数の解釈が「知覚の座」をめぐって競合するのである．

4. 錯聴を生み出す神経ネットワーク

錯聴は，知覚情報処理の脳内メカニズムを解明する上で貴重な情報源となる．今世紀に入って脳の非侵襲計測が急速に発展し，音声知覚時の脳活動を観察することが広く行われるようになった．しかし，「知覚そのもの」に対応する脳活動を捉えるのはさほど簡単ではない．仮に，音声信号 A に対して知覚 P が，音声信号 B に対して知覚 Q が生じるとしよう．音声信号 A と B とを聴取している時の脳活動がそれぞれ X, Y だったとしても，それは音声信号の物理特性（A, B）を反映したものか，それとも知覚内容（P, Q）を反映したものか区別できない．一方，知覚的修復のように，物理的に異なる音声 A と A'では知覚 P が生じ，音声 B と B'では知覚 Q が生じるとき，観察された脳活動が音声 A と A'に対しては X，音声 B と B'に対しては Y であれば，その脳活動は知覚内容（P, Q）を反映したものである可能性が高い．また，多義的知覚のように，同一の音声 A に対して知覚が P から P'へと変化する場合，脳活動がそれに伴って X, X'と変化するならば，その脳活動は物理特性ではなく知覚内容を反映したものである可能性が高い．つまり，錯聴における物理特性と知覚内容の乖離は，知覚の形成に関連する脳活動を特定する上で重要な手がかりとなるのである．

ここでは，VT を素材として，fMRI による脳機能計測を行った我々の研究を紹介する（Kondo & Kashino, 2007; Kashino & Kondo, 2012）．実験では，同一の単語（「バナナ」）を切れ目なく反復呈示し，知覚内容が変化したという聴取者の報告（ボタン押し）に同期した脳活動を検出した．その結果，脳内のいくつかの部位で有意な活動変化が認められた（ボタン押しに伴う運動成分は取り除いてある）（図 2）．

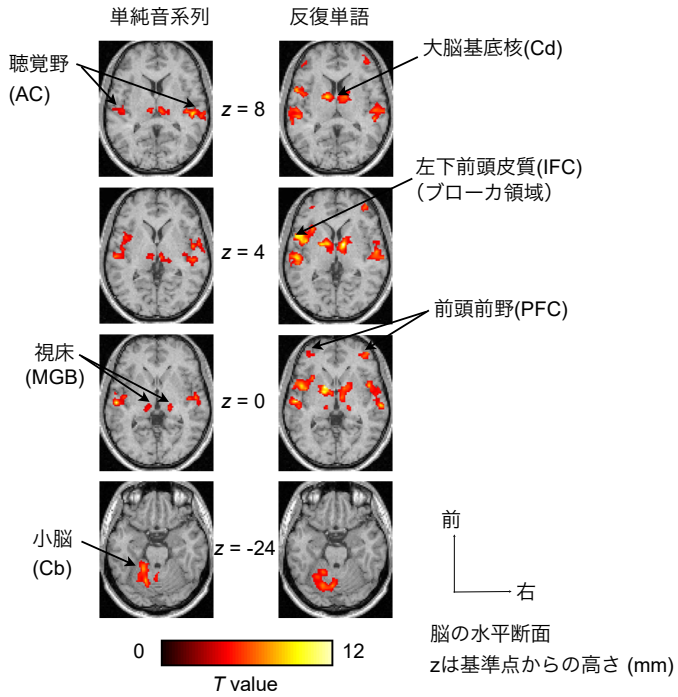


図 2: 単純な音系列（左）と反復単語（右）に対する知覚交替に同期した脳活動（各条件 12 人）．T value が大きいほど，その部位の知覚交替に伴う信号変化が顕著なことを示す．

まず，視床の内側膝状体（MGB），左後部島皮質（PIC），聴覚野（AC）を結ぶネットワークは，単純な音系列の多義的知覚を対象とした別実験（Kondo & Kashino, 2009）と共通に活動しており，基本的な音のグルーピングに関与していることが示唆された．

VT の実験のみで活動が見られたものに，左下前頭皮質（IFC），前頭前野（PFC），前部帯状皮質（ACC），大脳基底核の尾状核（Cd）があった．このうち，IFC の活動変化量は聴取者ごとの知覚交替の頻度と正の相関が見られた．一方，ACC の活動変化量は知覚交替の頻度と負の相関が見られた．知覚変化にとって，IFC がアクセル，ACC がブレーキの役割を果たしており，知覚交替の頻度は両者のバランスで決まるようである．IFC はいわゆるブローカ領域に相当し，特に弁蓋部は調音運動に関係する領域でもあるので，知覚の生成に調音運動のプロセスが関与するという「音声知覚の運動理論」（Liberman, et al., 1967）と親和性が高い. 音声の処理機構として最近有力視されている「二重経路モデル」（Hickok & Poeppel, 2007; Specht, 2014）では，聴覚野から上側頭溝，中・下側頭回後部を経て左 IFC 三角部に至る腹側経路は意味理解を司り，聴覚野から左頭頂側頭接合部を経て左運動前野，左 IFC 弁蓋部に至る背側経路は聴覚情報から運動情報への変換を司るとされる．VT で見られた活動は，この背側経路とも符合し，音声知覚への運動系の関与を示唆している．

さらに，VT の知覚交替に同期して Cd と MGB が連動しており，両者の連動性の強さは知覚交替の頻度が高い人ほど強いこともわかった．Cd は身体運動のタイミング制御，比較的長い時間スケールの複雑な時間構造の処理などに関与することが知られている（Zatorre, et al., 2007）．このネットワークは，音声の時間的分節に関与している可能性がある．

以上のように，VT における知覚の生成と選択には，脳内に広範に分散したいくつかの部位を結ぶネットワークが関与していることが示された．

5. 音声知覚の処理原理

音声知覚が脳内でどのように実現されているかを考える上で避けて通れない問題には次のようなものがある．(1)連続発声された音響信号が，いかにして離散的な言語的単位（音素など）へと分節化されるか，(2)音素環境や話速など，発声時の様々な変動要因の影響を受けた音響信号からいかに言語的単位が復号化されるか，(3)妨害音や伝送歪みなど，様々な外来変動要因に対して，いかに頑健な知覚を実現するか．これらの問題に対して，錯聴の特性分析や脳機能計測から重要なヒントが得られる．

知覚的修復，多義的知覚の両者とも，音声知覚が極めて能動的なプロセスであることを物語っている．耳に入力される音響信号は発話内容に関して完全な情報を与えるとは限らないが，ある程度の時間範囲に分散した特徴を統合することによって安定した知覚が実現されている．日常場面では，統合される情報は音響的なものに限らず，調音動作の映像のような視覚的なものや，文の構文や意味，その場の状況など，多様な情報が含まれうる．

能動的な処理を説明する一つの有力な枠組みは，予測符号化（predictive coding）という考え方である．感覚器に与えられる情報と，脳内に蓄えられた知識から，発話内容に関する予測（仮説）が生成される．その予測が，入力されてくる感覚情報と照合され，矛盾（予

測誤差）が少なければ予測が知覚として採用される．ある程度以上予測誤差が大きければ，その予測が棄却され，別の予測が生成される．このようなトップダウンの予測とボトムアップの感覚情報との照合が並列的，階層的に行われるという見方は，錯聴をはじめ，音声知覚の様々な側面をうまく説明することができる．

予測の一部は，調音運動の内部モデルから生成される可能性がある．この考え方の原型は運動理論として約半世紀前に提案されているが，今日なお議論が続いている．二重経路モデルにおける背側経路の役割も，音声知覚に必須なものであるか否か，決着がついていない．これらの点に関しては今後さらに検討が必要であろう．一つの可能性として，音声信号が劣化しているほど，運動系からの予測の貢献が大きくなると予想できる．

音声知覚の処理原理や神経メカニズムの探求は，この 10 年で飛躍的に進んでいる．その中で，錯聴現象は切れ味の鋭い道具を提供してきたし，今後もそうであろう．

参考文献

Bregman, A.S. (1990) *Auditory scene analysis: The perceptual organization of sound*. Cambridge: MIT Press, 1990.

Hickok, G., Poeppel, D. (2007) The cortical organization of speech processing. *Nature Reviews Neuroscience*. 8:393-402.

柏野牧夫 (1992) 「閉鎖区間の前後に分散する手がかりに基づく日本語語中閉鎖子音の知覚」『日本音響学会誌』48(2):76-86.

柏野牧夫 (2010) 『音のイリュージョン―知覚を生み出す脳の戦略』岩波科学ライブラリー 168, 東京：岩波書店.

柏野牧夫 (2012) 『空耳の科学―だまされる耳，聞き分ける脳』東京：ヤマハミュージックメディア.

Kashino, M., Kondo, H.M. (2012) Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 367(1591):977-987.

Kondo, H.M., Kashino, M. (2007) Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage.* 36(1):123-130.

Kondo, H.M., Kashino, M. (2009) Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *The Journal of Neuroscience*. 29:12695-12701.

Liberman, A.M., Cooper, F.S., et al. (1967) Perception of the speech code. *Psychological Review*. 74(6):431-461.

Specht, K. (2014) Neuronal basis of speech comprehension. *Hearing Research.* 307: 121-135.

Warren, R.M. (2008) *Auditory perception: An analysis and synthesis*. Cambridge: Cambridge University Press.

Zatorre, R.J., Chen, J.L., Penhune, V.B. (2007) When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*. 8:547-558.

公開講演2（PL2）

乳幼児における聴知覚の発達

出口　利定（東京学芸大学）

１．はじめに

乳児は始語が出現する以前の生後１年間に、音声知覚能力の著しい発達を遂げることが知られている。この場合の発達とは、単に音響刺激に対する聴覚末梢系の分析能力の向上という意味でなく、音声を言語として処理する過程である。したがって、乳幼児における音声知覚研究は、言語獲得に過程における音声言語知覚能力の生得的特性と、その知覚特性が母語環境の影響のもとでどのように母語に特異的に適合/発達していくかを解明するうえで重要である。

1970 年代後半からの乳児の音声知覚に関する実験的研究から、多くの興味深い知見が得られている。当初は、音韻弁別能力に関する研究が行われ、それらの成果から、乳児は早期には、その母語にかかわらず、世界中の言語に存在するほぼ全ての音韻（有限ではある）を聴覚的に弁別する能力を有し、そのユニバーサルな音韻知覚能力は、月例が高くなるにつれ、母語特有の音韻体系に規定された、より成人に近い知覚特性を示すようになることが明らかになった。

そしてそのユニバーサルな知覚特性から母語の影響を受けた知覚特性への変化の時期は、母音知覚に関しては、生後６ヶ月前後、子音知覚に関しては生後１０ヶ月前後を境に、母語には存在しない非母語の音韻対立の弁別能力は低下する。すなわち、母語の音韻体系のカテゴリーへの同化能力の発達を意味し、音声処理がより母語に適合してきたことを示すと考えられる。

また、乳児は音韻レベルの違いよりも、むしろ音声の韻律的な情報に、かなり早期から注意を向けていることが知られている。従来より、音声言語における韻律の変化は話者の感性情報を伝えるのみでなく、発話の意味的・統語的な分節位置を示す手がかりとなることが知られている。このような観点から、乳児における韻律情報の知覚能力の発達と、節、句、単語などの音声の構造単位のセグメンテーション能力との関連が精力的に研究されている。

乳幼児の音声知覚の発達を明らかにするためには、乳幼児の音声刺激に対する身体的な反応・行動を指標として研究が進められてきた。知覚機構の発達・変化の背景には、音声処理に対する神経回路の発達・再構成の関与が考えられることから、音声獲得、言語獲得の研究においても近年では脳活動の計測技術の飛躍的な進歩により、脳科学的研究が多くなり新しい知見が得られている。そういう時代的な背景にあっても、音声知覚研究における乳幼児の行動を指標とした研究の重要性、困難性、楽しさ？について、既に報告済のややノスタルジックな発表である。なお、紹介する研究は、桐谷滋（東京大学）、林安紀子、加藤ふみ（東京学芸大学）との共同研究である。

２．　吸啜行動を指標とした母語-非母語の韻律情報の識別

乳児の吸啜行動を指標とした知覚実験は多くあるが、吸啜そのものの性質、特性、乳児の状態（月齢、空腹度、母乳・人工乳・混合、覚醒度等）との関係は明らかではなく、最適な吸啜行動を得るための条件を決定することが重要である。図１に、吸啜行動が始まってから終了（入眠）までの吸啜行動の実際の状態を示す。

吸啜区間と休止区間が交互に出現し、リズミカルな運動を示している。この吸啜区間と休止区間が交互に出現する、つまり吸啜区間のみが最後まで連続することはなく、この現象は他の霊長類にはみられずヒト特有であるという指摘は、母子コミュニケーション上、非常に興味深い。

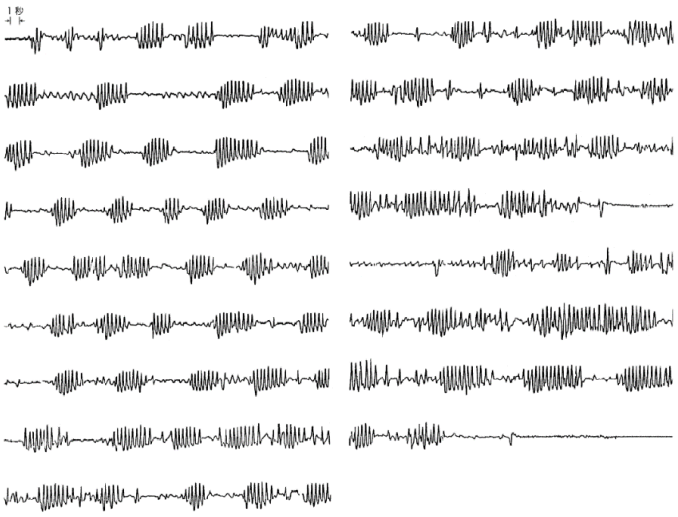


図１．生後 23 日の乳児の吸啜

図２に、母語（日本語）と非母語（英語）に対する吸啜回数の変化（平均）を示す。

まず日本語：J（英語：E）を聞かせながら吸啜開始から最初の 1 分間の吸啜回数をベースラインとし、1 分毎の吸啜回数が 25%以上、2 回連続して観測されたら慣化とみなし、音声刺激を英語：E（日本語：J）に切り替えた。日本語から英語（J→E）、英語から日本語（E→J）ともに刺激変換後に吸啜回数の上昇がみられた。乳児が言語刺激の変化を弁別し、かつ母語に対して親密性を持っている、あるいは母語に新奇性を強めるものがあるのではないかと推測される。

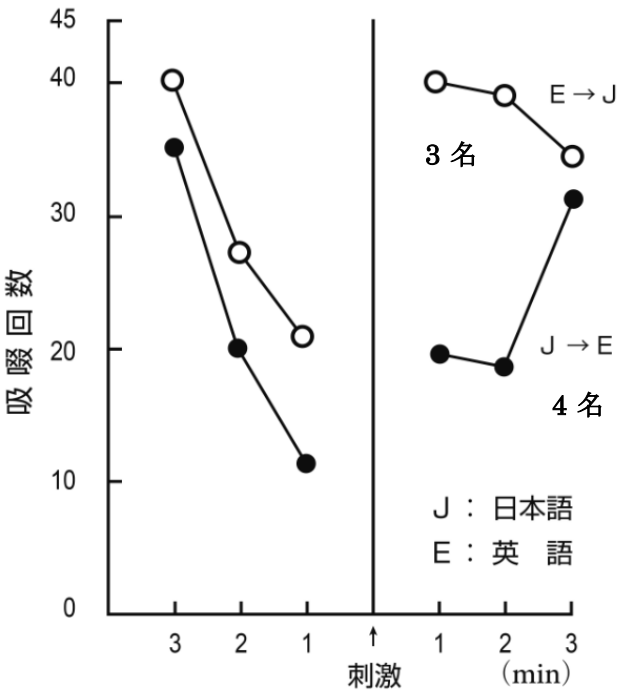


図２．母語・非母語への変化による吸啜回数の変動

３．母音知覚における正規化能力の発達

図３は、成人男性、成人女性、子どもによって発声された日本語５母音の第１ホルマント、第２ホルマントの平均的な分布を示すよく知られた図である（千葉、1990）。五角形はほぼ相似形を保っているものの位置が異なっている。ホルマント周波数の一様な移動は、計算機モデルで言えば声帯から唇までの長さ、即ち声道長に比例する。

基本周波数、ホルマント周波数は異なっても、同一の母音として知覚しているこの現象は、音声知覚の恒常性、あるいは声道長の正規化と呼ばれ、音声知覚にとっては非常に重要である。

この正規化能力について、幼児・乳児の母音知覚についても成人同様に正規化の過程が作用しているかについて、発達的な観点から検討した。

刺激母音は成人男性、成人女性、子どもを含む多数話者の自発発声サンプルから求めたパラメータを主として用い（Fujisaki,1968）、F1、F2 周波数を連続的に変化させ、/o/から/a/へ推移する/o－a/系列 11 音と、/u/から/e/へ推移する/u－e/系列 13 音の 2 系列である。また、各系列に対して、2 種類の基本周波数 100Hz,220Hz を設定した（図４）。

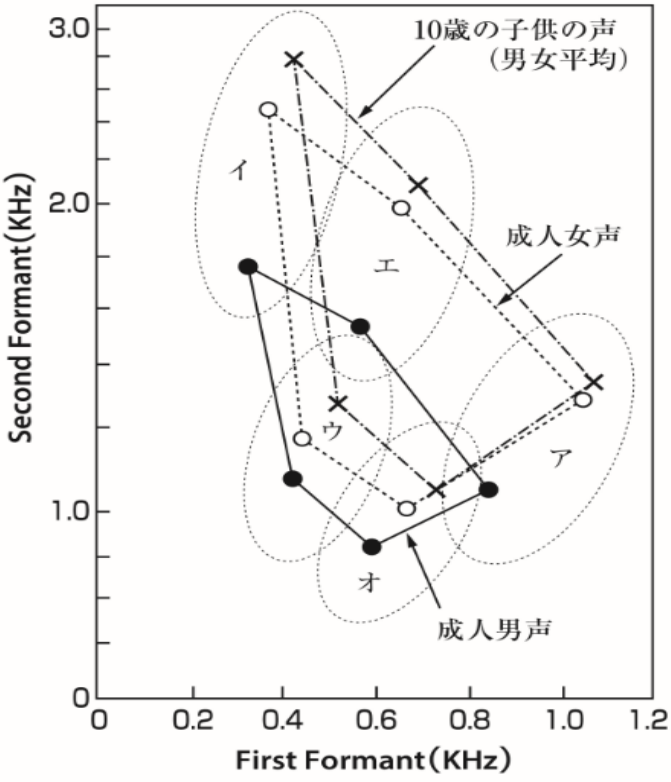
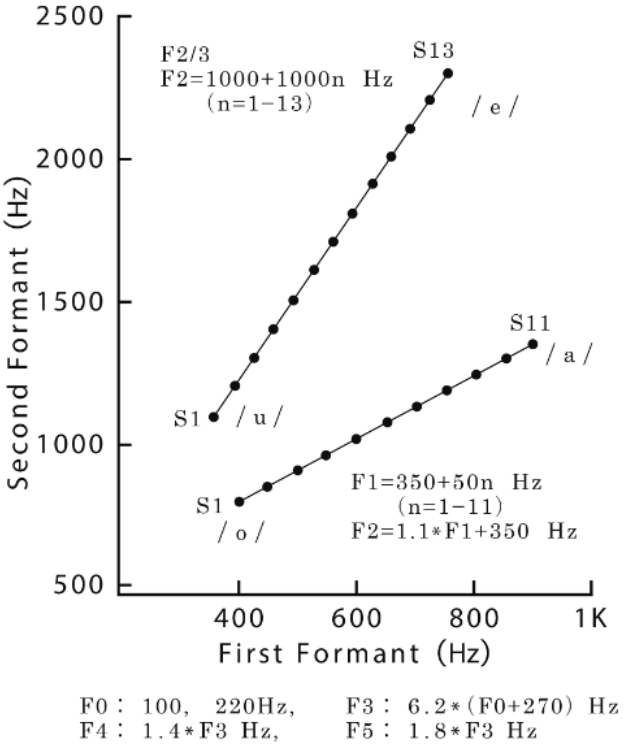


図３．成人男性・女性、子どもの母音発声における F1,F2 の位置



F0： 100, 220Hz, F3： 6.2*(F0+270) Hz
F4： 1.4*F3 Hz, F5： 1.8*F3 Hz

図４．刺激音

図５は成人 12 名、3 歳児 5 名の/u－e/系列における聴取結果である。横軸は図４の刺激音番号に対応し、各刺激音に対して/e/と判断した割合を示した。

被験者ごとに各系列における識別カーブから/e/あるいは/u/と判断した割合の推移を最尤法によって正規分布として近似し、平均値と標準偏差を求め、平均値は判断境界の刺激音番号を、標準偏差は判断の精度を示すものとした。

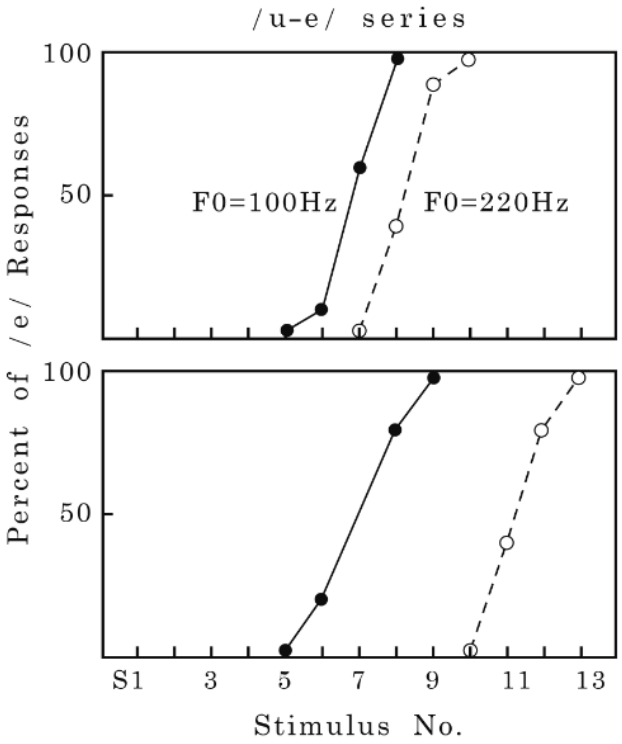


図５．成人（上）、3 歳児（下）の聴取例

図６に/u－e/系列、/o－a/系列における判断境界を示す刺激音番号の平均と範囲を、被験者グループ毎に基本周波数別に示した。

基本周波数の変化に伴う判断境界の移行量は、いずれの系列も幼児の方が成人よりも多い傾向にあった。

幼児は成人に比較して、母音知覚において基本周波数および高次ホルマント周波数の変動による影響を受け易いことが示された。

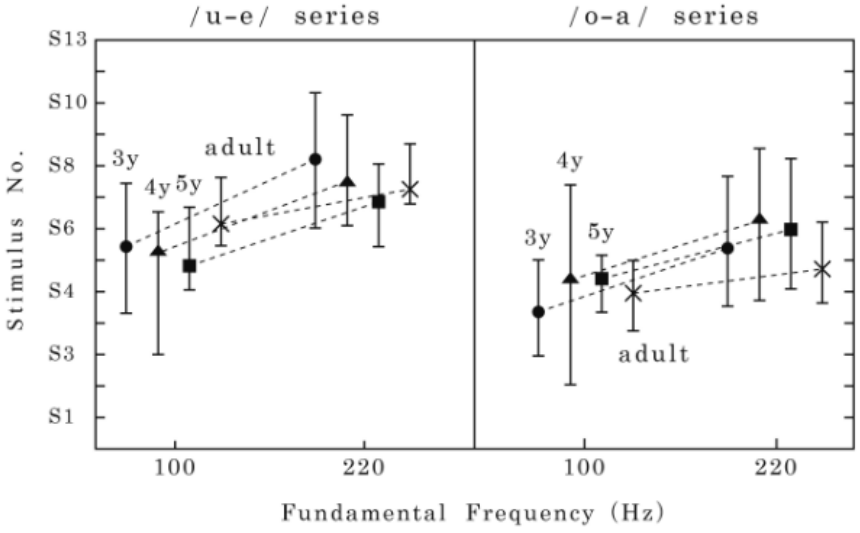


図６．基本周波数毎の判断境界の平均と判断境界の移行量

４．乳児における正規化能力

月齢が 5 カ月から 7 カ月までの乳児 5 名を対象に、乳児を対象とした音声知覚の手法として多く採用されている head-turn 法を用いて、/o－a/系列について母音聴取における正規化能力を調べた。

head-turn 法の概略は、一定間隔で、ある同一の刺激音あるいは同一カテゴリ内の刺激音のセットを提示し続け（background 刺激）、時に新奇刺激（comparison 刺激）を提示する状況で新奇刺激の出現に気付いたときのみ音源を振り向く反応（head-turn）、および background 刺激には振り向かない反応を、まずは予備練習によって乳児に習得させる。この head-turn 反応の生起率を、1 対の刺激音（または刺激セット）の弁別判断の指標とする方法である。この方法は、生後約 5 か月から約 8 ヶ月までと、適用月齢が限られてはいるが、head-turn 反応を乳児が獲得すれば、統制群の必要がなく、個人内データで音声知覚について議論できることが最大の長所である。しかしながら head-turn 反応の獲得には熟達したテクニックを必要とすることから、多くの時間を要し、乳児の扱いに慣れた実験助手の育成も必要である。

予備練習の段階（training phase）では、background 刺激（S11：典型的な/a/）に対しては head-turn はせず、comparison 刺激（S1：典型的な/o/）に対してのみ head-turn するという反応を確実に獲得させる（図 7）。その際、正しい反応に対しては、視覚・聴覚的報酬としてブラックボックス内の玩具に照明を当て作動させた（図 8）。本実験（experimental phase）では、comparison 刺激として S1 から S10 の刺激音（intermediate vowel）を提示した。

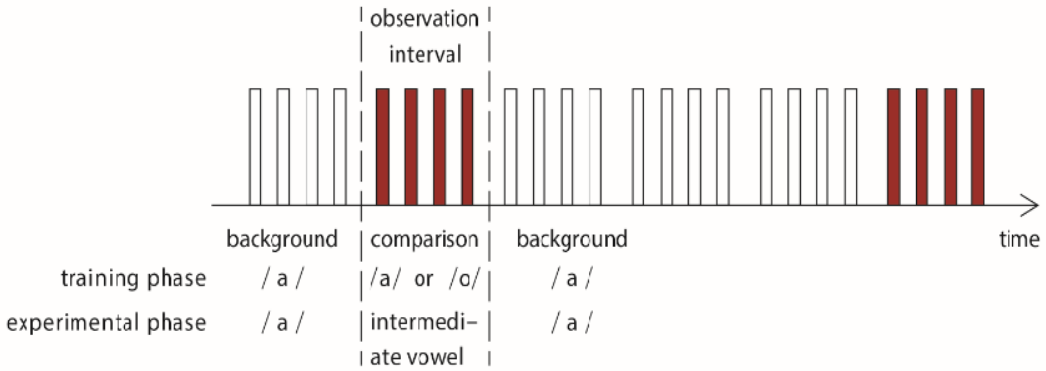


図 7．Head-turn 法における刺激音の提示

experimental phase では、comparison刺激としてS1からS10をそれぞれ5回提示し、head-turn生起率を求め弁別判断曲線を得た（図９）。

background 刺激から音響的に離れた comparison 刺激ほど head-turn反応が増加し、基本周波数及び高次ホルマント周波数の影響を受けて弁別判断境界が移行することが判った（図 10）。

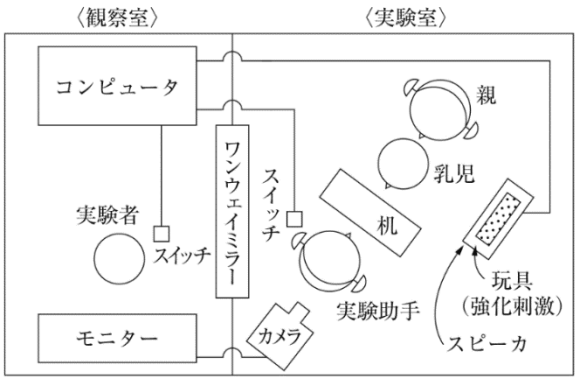


図８．実験室の配置

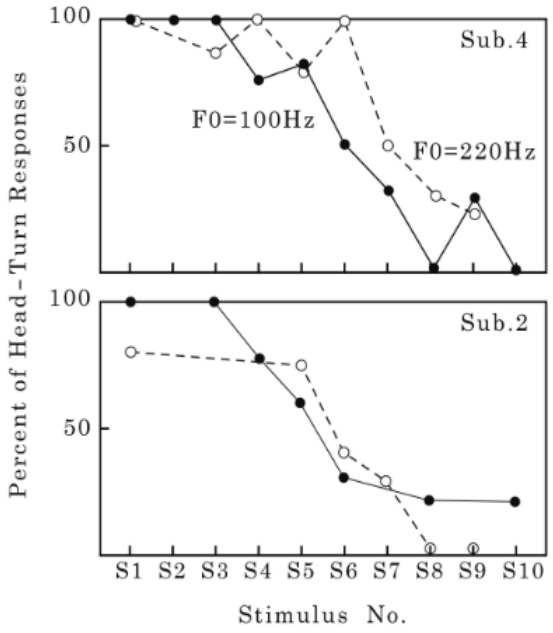


図９．Head-turn 反応による乳児の弁別曲線例

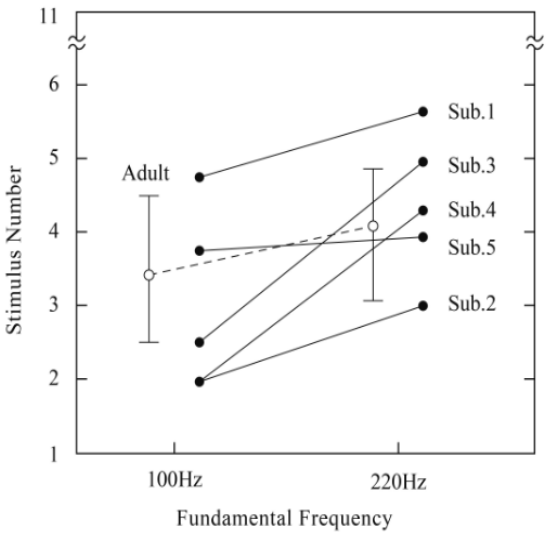


図 10．乳児５名の基本周波数毎の判断境界の平均と判断境界の移行量

吸啜反応（行動）、head-turn 法を用いた、乳幼児の母音知覚実験について述べたが、特に乳児を対象とした音声知覚実験手法については、より洗練された手法の開発が必要である。また得られた知見を裏付ける脳科学的な解明が望まれる。いかにしてヒトは言語を獲得していくのか？身近にいるボンヤリした我が孫でも、いざ言語獲得という観点から眺めると神々しくみえ、その計算過程を教えてくれと言いたくなる。

A1

高知市方言の上昇遅れと下降早まり

高山 林太郎（東京大学大学院）
takayama_rintaro@nifty.com

1. はじめに—“なまけ”について

中井幸比古（1997）『高知市方言アクセント小辞典』と同じ話者（下原瑞惠氏）を調査し，同辞典に記されている音韻現象の詳細な記述を目指した。小辞典には「低起式の上昇の 1 拍遅れ」（例：ウ[サギ～ウサ[ギ（兎），ア[ルイチュ]ー～アルイ[チュ]ー（歩いとる））と「下げ核の下降の 1 拍早まり」（例：[アカル]イ～[アカ]ルイ（明），[キルま]い～[キル]まい（着））の現象が具体例として記述されている。これらの現象を下原氏は“なまけ”（造語）と呼び，古い世代の発音に対する若い世代の発音であり，自分にとってより楽な発音であって，使用頻度が高い発音であるとする。他方で，“なまけ”を起こさない発音は，古い世代の発音であり，使用頻度が低いが自分も使用することはあり，特に強調を表す場合には選好して使用する発音であるとする（また逆に，“なまけ”を起こした発音を，強調を表す為に用いるケースも中にはある）。音韻現象としては「低起式の上昇の 1 拍遅れ」と「下げ核の下降の 1 拍早まり」は全くの別物であるが，話者の心内辞書という共時的な体系内において占める通時的な地位が同様であることから下原氏はこれらをいずれも“なまけ”と呼ぶのであって，これらの音韻現象の通時的な位置付けを“なまけ”という造語は示している。本稿では下原氏が“なまけ”であると確かに内省した現象だけを取り扱っている。

2. 低起式の上昇の 1 拍遅れについて

「低起式の上昇の 1 拍遅れ」については，モーラ音素を含む語について見ると「イッ[コ（一個），ア[ンキ～アン[キ（暗記），ギョ[ーギ～ギョー[ギ（行儀），サ[イショ～サイ[ショ（最初）」のような上昇位置が下原氏にはありうる。「ア[ンキ，ギョ[ーギ，サ[イショ」も“なまけ”以前の形として使うことがあるという点が小辞典の記述と一部異なる。また，「ア[ルイチュ]ー～アルイ[チュ]ー（歩），ハ[イッチュ]ー～ハイッ[チュ]ー（入），カ[カエチュ]ー～カカエ[チュ]ー（抱）」のように「1 拍遅れ」の「拍」がモーラでなく音節または音節相当のものに該当するケースも少数ながら存在する（※小辞典には「カ[カエチュ]ー～カカ[エチュ]ー」とあるが，その発音も可能であろうと考える）。モーラ数の多い複合語などの場合には複合語内部境界などにまで“なまけ”が進みうる点は小辞典の記述通りである。「低起式の上昇の 1 拍遅れ」については他に特に述べるべきことはない。

3. 下げ核の下降の 1 拍早まりについて

「下げ核の下降の 1 拍早まり」については，小辞典の記述より更に詳しいことが判明したので，本稿ではこの点を主に記述する。形容詞・情態副詞など（例：[アカル]イ～[アカ]ルイ（明），(1) [カルガ]ルと～[カル]ガルと（軽々），(2) [イジ]イジと～[イ]ジイジと，(3) [オキた]い～[オキ]たい（置），(4) [キルま]い～[キル]まい（着），(5) [キリャ]ー～[キ]リャ）で

起こる現象であり，形容詞について規則的に起こる（4 拍以上の高起式の n 拍形容詞に Hn-1 型と Hn-2 型が併用され，5 拍以上の低起式の n 拍形容詞に Ln-1 型と Ln-2 型が併用され，Hn-2 型や Ln-2 型が“なまけ”に当たる）ことは小辞典の記載通りであるからその点について本稿ではこれ以上は扱わないこととする。

そのほか，「順接」（高く入れば高く付き，低く入れば低く付く）の有核の機能語が「低接」（必ず低く付く）に変化すること（例：エ[ー の]わ～エ[ー]のわ（良；伝統的な形式体言「が」とは別に共通語的な形式体言「の」も使用され，変化を被ったと考えられる），(6) [ニワの]わ～[ニワ]のわ（庭），(7) [ニワよ]り～[ニワ]より，(8) [ニワこ]そ～[ニワ]こそ，(9) [ニワな]ら～[ニワ]なら，(10) [ニワま]で～[ニワ]まで，(11) [ニワら]ー～[ニワ]らー，(12) [ニワさ]え～[ニワ]さえ，(13) [ニワし]か～[ニワ]しか，(14) [ニワや]ら～[ニワ]やら，(15) [ニワな]り～[ニワ]なり，(16) [ニワで]も～[ニワ]でも，(17) [キルけ]んど～[キル]けんど（着），(18) [キルま]で～[キル]まで，(19) [ザツな]ら～[ザツ]なら（雑），(20) [ザツで]も～[ザツ]でも）も下原氏曰く“なまけ”であり，「下げ核の下降の 1 拍早まり」という大きな枠組みの中に取り込まれるものである。なお伝統的な形式体言「が」は「高接」（直前に下げ核があればそれを消去してでも高く付く）であり，例えば「[シ]ロイ（白）」に「が]わ」が付くと「[シロイが]わ」となるもので，有核ではあるが順接ではないので「下げ核の下降の 1 拍早まり」は起こさない。以下では番号 (1)～(20) を振った現象を個別に見ていく。

3.1. 「[カルガ]ルと～[カル]ガルと（軽々）」等について（情態副詞）

情態副詞であり，高山林太郎（2013）で扱った「*アカア[カ]と」の類に該当する。京都では平安末期の段階で早くも「*[ピ]カピカと」の類に合流済みであるが，周辺の中央式の方言ではそうではない。高知市では「*アカア[カ]と（平安末期相当）>*[アカ]アカと（室町相当）>[アカア]カと>[アカ]アカと」のように，一旦下げ核の位置が 3 拍目に移動したあと 2 拍目に巻き戻ったと見られ（※東京や岡山市では「ア[カア]カと」であり，3 拍目に移動したままになっている），下原氏が 2 種類の形を内省するのは通時的変化の前後の形をそのまま記憶しているものと考えられる。語例：[アオア]オと～[アオ]アオと，[アカア]カと～[アカ]アカと，[クログ]ロと～[クロ]グロと，[シロジ]ロと～[シロ]ジロと，…。

3.2. 「[イジ]イジと～[イ]ジイジと」等について（情態副詞）

情態副詞であり，高山（2013）で扱った「*[ピ]カ[ピ]カと～*[ピ]カピカと」の類に該当する。高知市や周辺部の中央式の方言では通常形の異音としての 2 単位形「[ピ]カ[ピ]カと」が存在することを高山（2013）で述べたが，下原氏の場合，2 単位形は強調形である。1 単位形について考えると，「*[ピ]カピカと（平安末期相当）>*[ピカ]ピカと>[ピ]カピカと」のように，一旦下げ核の位置が 2 拍目に移動したあと 1 拍目に巻き戻ったと見られ（※東京では 1 拍目が無声化する環境においてのみ巻き戻らず，「ピ[カ]ピカと，[キ]ラキラと」であり，岡山市では一切巻き戻らず，「ピ[カ]ピカと，キ[ラ]キラと」である），下原氏が 2 種類の形を内省するのは通時的変化の前後の形をそのまま記憶しているものと考えられる（※

なお上記で「*[ピカ]ピカと」が再構形になっているのは，この語例に関しては回答が得られなかった為であり，下原氏の記憶にあれば古形が回答されるが無ければ回答されず，下原氏は機械的に作り出した形としてこれらの形を回答している訳ではない)。語例：[イジ]イジと～[イ]ジイジと，[イラ]イラと～[イ]ライラと，[ウカ]ウカと～[ウ]カウカと，[ウキ]ウキと～[ウ]キウキと，…。

3.3. 「[オキた]い～[オキ]たい（置）」等について（動詞文節）

上述した形容詞の規則と同じである。語例：[キた]い（着），[シた]い（為），[ミた]い（見），[キた]い（来），[オキた]い～[オキ]たい（置），[フミた]い～[フミ]たい（踏），[カキた]い～[カキ]たい（書），[ノミた]い～[ノミ]たい（飲），[オリた]い～[オリ]たい（居），[アケた]い～[アケ]たい（開），[カエた]い～[カエ]たい（変），[タテた]い～[タテ]たい（建），[ナゲた]い～[ナゲ]たい（投），[キザミた]い～[キザミ]たい（刻），[ツナギた]い～[ツナギ]たい（繋），[コガシた]い～[コガシ]たい（焦），[サワギた]い～[サワギ]たい（騒），ア[ルキた]い～ア[ルキ]たい（歩），ハ[イリた]い～ハ[イリ]たい（入），[シラセた]い～[シラセ]たい（知），[ハジメた]い～[ハジメ]たい（始），[アズケた]い～[アズケ]たい（預），[アツメた]い～[アツメ]たい（集），カ[カエた]い～カ[カエ]たい（抱）。

3.4. 「[キルま]い～[キル]まい（着）」等について（動詞文節）

形容詞の規則とほぼ同じであり，n-1 拍目に下げ核がある場合に変化が起こるが，低起式でも 4 拍語から変化が起きている。語例：[キルま]い～[キル]まい，[スルま]い～[スル]まい，ミ[ルま]い～ミ[ル]まい，ク[ルま]い～ク[ル]まい，[オクま]い～[オク]まい，[フムま]い～[フム]まい，カ[クま]い～カ[ク]まい，ノ[ムま]い～ノ[ム]まい，[オルま]い～[オル]まい，[アケルま]い～[アケル]まい，[カエルま]い～[カエル]まい，[タ]テルまい，[ナ]ゲルまい，[キザムま]い～[キザム]まい，[ツナグま]い～[ツナグ]まい，[コ]ガスまい，[サ]ワグまい，ア[ルクま]い～ア[ルク]まい，ハ[イルま]い～ハ[イル]まい，[シラセルま]い～[シラセル]まい，[ハジメルま]い～[ハジメル]まい，[アズ]ケルまい，[アツ]メルまい，カ[カエルま]い～カ[カエル]まい。

3.5. 「[キリャ]ー～[キ]リャ（着）」等について（動詞）

n-1 拍目に下げ核がある場合に変化が起こり，その際に長音が消滅する場合がある。語例：[キリャ]ー～[キ]リャ，[スリャ]ー～[ス]リャ，ミ[リャ]ー～[ミ]リャ（※低起式が高起式に変わるケースもある），ク[リャ]ー～[ク]リャ，[オキャ]ー～[オ]キャ，[フミャ]ー～[フ]ミャ，カ[キャ]ー，ノ[ミャ]ー，[オリャ]ー～[オ]リャ(ー)，[アケリャ]ー～[アケ]リャー，[カエリャ]ー～[カエ]リャ(ー)，[タ]テリャー～タ[テ]リャ（※類推変化と見られる），[ナ]ゲリャー～ナ[ゲ]リャ，[キザミャ]ー～[キザ]ミャ(ー)，[ツナギャ]ー～[ツナ]ギャ，[コガシャ]ー～[コガ]シャ(ー)～[コ]ガシャ(ー)，[サ]ワギャー～[サ]ワギャ，ア[ルキャ]ー～ア[ル]キャ，ハ[イリャ]ー～ハ[イ]リャ，[シラセリャ]ー～[シラセ]リャ(ー)～[シラ]セリャ(ー)，[ハジメリャ]ー～[ハジメ]リャ(ー)～[ハジ]メリャ，[アズケリャ]ー～[アズケ]リャ～[アズ]ケリャ(ー)～[ア]ズケリ

ャー, [アツメリャ]ー〜[アツメ]リャ(ー)〜[アツ]メリャ(ー)〜[ア]ツメリャ(ー), カ[カエリャ]ー〜カ[カエ]リャ(ー)〜カ[カ]エリャ(ー)。

3.6. 「[ニワの]わ〜[ニワ]のわ（庭）」等について（名詞文節）

形容詞の規則と同じであり，n-1 拍目に下げ核がある場合に変化が起こる。語例：[血の]わ，[巣の]わ，[名]のわ，[歯]のわ，手[の]わ，目[の]わ，[ニワの]わ〜[ニワ]のわ（庭），[ミゾの]わ〜[ミゾ]のわ（溝），[ゴマの]わ〜[ゴマ]のわ（胡麻），[ヤ]マのわ（山），[イ]シのわ（石），ハ[リの]わ（針），ア[メ]のわ〜ア[メの]わ（雨），[コトリの]わ〜[コトリ]のわ（小鳥），[トコロの]わ〜[トコロ]のわ（所），[チ]カラのわ（力），[イ]ノチのわ（命），[ア]サヒのわ（朝日），[アタ]マのわ（頭），[ムス]メのわ（娘），ツ[バ]キのわ（椿），ク[ス]リのわ（薬），ス[ズメの]わ〜ス[ズメ]のわ（雀）。

3.7. 「[ニワよ]り〜[ニワ]より（庭）」等について（名詞文節）

形容詞の規則とほぼ同じであり，n-1 拍目に下げ核がある場合に変化が起こるが，3 拍文節から変化が起きている。語例：[血よ]り〜[血]より，[巣よ]り〜[巣]より，[名よ]り〜[名]より，[歯よ]り〜[歯]より，手[よ]り，目[よ]り，[ニワよ]り〜[ニワ]より，[ミゾよ]り〜[ミゾ]より，[ゴマよ]り〜[ゴマ]より，[ヤ]マより，[イ]シより，ハ[リよ]り〜ハ[リ]より，ア[メよ]り〜ア[メ]より，[コトリよ]り〜[コトリ]より，[トコロよ]り〜[トコロ]より，[チ]カラより，[イ]ノチより，[ア]サヒより，[アタ]マより，[ムス]メより，ツ[バ]キより，ク[ス]リより，ス[ズメよ]り〜ス[ズメ]より。

3.8. 「[ニワこ]そ〜[ニワ]こそ（庭）」等について（名詞文節）

規則は同上。語例も同上（「より」を「こそ」に変えただけ）。

3.9. 「[ニワな]ら〜[ニワ]なら（庭）」等について（名詞文節）

規則は同上。語例：[血な]ら〜[血]なら，[巣な]ら〜[巣]なら，[名]なら，[歯]なら，手(ー)[な]ら，目(ー)[な]ら，[ニワな]ら〜[ニワ]なら，[ミゾな]ら〜[ミゾ]なら，[ゴマな]ら〜[ゴマ]なら，[ヤ]マなら，[イ]シなら，ハ[リな]ら〜ハ[リ]なら，ア[メな]ら〜ア[メ]なら，[コトリな]ら〜[コトリ]なら，[トコロな]ら〜[トコロ]なら，[チ]カラなら，[イ]ノチなら，[ア]サヒなら，[アタ]マなら，[ムス]メなら，ツ[バ]キなら，ク[ス]リなら，ス[ズメな]ら〜ス[ズメ]なら。

3.10. 「[ニワま]で〜[ニワ]まで（庭）」等について（名詞文節）

形容詞の規則とほぼ同じであり，n-1 拍目に下げ核がある場合に変化が起こるが，低起式でも 4 拍文節から変化が起きている。語例：[血]まで，[巣]まで，[名]まで，[歯]まで，手(ー)[ま]で，目(ー)[ま]で，[ニワま]で〜[ニワ]まで，[ミゾま]で〜[ミゾ]まで，[ゴマま]で〜[ゴマ]まで，[ヤ]マまで，[イ]シまで，ハ[リま]で〜ハ[リ]まで，ア[メま]で〜ア[メ]まで，[コトリま]で〜[コトリ]まで，[トコロま]で〜[トコロ]まで，[チ]カラまで，[イ]ノチまで，[ア]サヒまで，[アタ]マまで，[ムス]メまで，ツ[バ]キまで，ク[ス]リまで，ス[ズメま]で〜ス[ズメ]ま

で。

3.11. 「[ニワら]ー〜[ニワ]らー（庭）」等について（名詞文節）

「らー」は「なんか・たち」の意である。形容詞の規則とほぼ同じであり，n-1 拍目に下げ核がある場合に変化が起こるが，3 拍文節から変化が起きている。語例：[血ら]ー〜[血]らー，[巣ら]ー〜[巣]らー，[名ら]ー〜[名]らー，[歯ら]ー〜[歯]らー，手(ー)[ら]ー，目(ー)[ら]ー，[ニワら]ー〜[ニワ]らー，[ミゾら]ー〜[ミゾ]らー，[ゴマら]ー〜[ゴマ]らー，[ヤ]マらー，[イ]シらー，ハ[リら]ー〜ハ[リ]らー，ア[メら]ー〜ア[メ]らー，[コトリら]ー〜[コトリ]らー，[トコロら]ー〜[トコロ]らー，[チ]カラらー，[イ]ノチらー，[ア]サヒらー，[アタ]マらー，[ムス]メらー，ツ[バ]キらー，ク[ス]リらー，ス[ズメら]ー〜ス[ズメ]らー。

3.12. 「[ニワさ]え〜[ニワ]さえ（庭）」等について（名詞文節）

規則は同上。語例も同上（「らー」を「さえ」に変えただけ）。

3.13. 「[ニワし]か〜[ニワ]しか（庭）」等について（名詞文節）

規則は同上。語例：[血し]か〜[血]しか，[巣し]か〜[巣]しか，[名]しか，[歯]しか，手(ー)[し]か，目(ー)[し]か，[ニワし]か〜[ニワ]しか，[ミゾし]か〜[ミゾ]しか，[ゴマし]か〜[ゴマ]しか，[ヤ]マしか，[イ]シしか，ハ[リし]か〜ハ[リ]しか，ア[メし]か〜ア[メ]しか，[コトリし]か〜[コトリ]しか，[トコロし]か〜[トコロ]しか，[チ]カラしか，[イ]ノチしか，[ア]サヒしか，[アタ]マしか，[ムス]メしか，ツ[バ]キしか，ク[ス]リしか，ス[ズメし]か〜ス[ズメ]しか。

3.14. 「[ニワや]ら〜[ニワ]やら（庭）」等について（名詞文節）

規則は同上。語例も同上（「しか」を「やら」に変えただけ）。

3.15. 「[ニワな]り〜[ニワ]なり（庭）」等について（名詞文節）

規則は同上。語例：[血な]り〜[血]なり，[巣な]り〜[巣]なり，[名な]り〜[名]なり，[歯な]り〜[歯]なり，手(ー)[な]り，目(ー)[な]り，[ニワな]り〜[ニワ]なり，[ミゾな]り〜[ミゾ]なり，[ゴマな]り〜[ゴマ]なり，[ヤ]マなり，[イ]シなり，ハ[リな]り〜ハ[リ]なり，ア[メな]り〜ア[メ]なり，[コトリな]り〜[コトリ]なり，[トコロな]り〜[トコロ]なり，[チ]カラなり，[イ]ノチなり，[ア]サヒなり，[アタ]マなり，[ムス]メなり，ツ[バ]キなり，ク[ス]リなり，ス[ズメな]り〜ス[ズメ]なり。

3.16. 「[ニワで]も〜[ニワ]でも（庭）」等について（名詞文節）

規則は同上。語例も同上（「なり」を「でも」に変えただけ）。

3.17. 「[キルけ]んど〜[キル]けんど（着）」等について（動詞文節）

n-2 拍目に下げ核がある場合に変化が起こる。語例：[キルけ]んど〜[キル]けんど，[スルけ]んど〜[スル]けんど，ミ[ルけ]んど〜ミ[ル]けんど，ク[ルけ]んど〜ク[ル]けんど，[オクけ]んど〜[オク]けんど，[フムけ]んど〜[フム]けんど，カ[クけ]んど〜カ[ク]けんど，ノ[ムけ]んど〜ノ[ム]けんど，[オ]ルけんど，[アケルけ]んど〜[アケル]けんど，[カエルけ]んど〜

[カエル]けんど，[タ]テルけんど，[ナ]ゲルけんど，[キザムけ]んど～[キザム]けんど，[ツナグけ]んど～[ツナグ]けんど，[コ]ガスけんど，[サ]ワグけんど，ア[ルクけ]んど～ア[ルク]けんど，ハ[イルけ]んど～ハ[イル]けんど，[シラセルけ]んど～[シラセル]けんど，[ハジメルけ]んど～[ハジメル]けんど，[アズ]ケルけんど，[アツ]メルけんど，カ[カエルけ]んど～カ[カエル]けんど。

3.18. 「[キルま]で～[キル]まで（着）」等について（動詞文節）

形容詞の規則とほぼ同じであり，n-1 拍目に下げ核がある場合に変化が起こるが，低起式でも 4 拍文節から変化が起きている。語例は同上（「けんど」を「まで」に変えただけ）。

3.19. 「[ザツな]ら～[ザツ]なら（雑）」等について（形容動詞文節）

規則は同上。語例：[ザツな]ら～[ザツ]なら（雑），[ス]キなら（好），イ[キな]ら～イ[キ]なら（粋），ウ[ブな]ら～ウ[ブ]なら（初），[イビツな]ら～[イビツ]なら（歪），[コマカな]ら～[コマ]カなら～[コマ]カなら（細），[オ]ロカなら（愚），ガ[サツな]ら～ガ[サツ]なら，ア[コ]ギなら（阿漕）。

3.20. 「[ザツで]も～[ザツ]でも（雑）」等について（形容動詞文節）

規則は同上。語例：[ザツで]も～[ザツ]でも，[ス]キでも，イ[キで]も，ウ[ブで]も～ウ[ブ]でも，[イビツで]も～[イビツ]でも，[コマ]カでも，[オ]ロカでも，ガ[サツで]も～ガ[サツ]でも，ア[コ]ギでも。

4. おわりに

3 節で見たように，（1）名詞文節・動詞文節・形容詞・形容動詞文節・情態副詞という幅広い内容語において「下げ核の下降の 1 拍早まり」が観察されること，また，（2）だいたい形容詞に見られるのと同じような「下げ核の下降の 1 拍早まり」が，元が順接の有核の機能語において各所で見られ，その結果として新しい発音においては低接に変化していること，の 2 点が分かった。（2）は高知市方言の機能語を扱う上で注意すべき点である。

参考文献

高山林太郎 (2013) 「日本語諸方言の四モーラ畳語を比較する試み」『東京大学言語学論集』34, 143-183. 東京：東京大学言語学研究室

中井幸比古 (1997) 『高知市方言アクセント小辞典』神戸：私家版

中井幸比古 (2002) 『京阪系アクセント辞典』東京：勉誠出版

A2

三重県鈴鹿市における 3 拍語のアクセント

竹内 はるか(國學院大學大学院)
y124209@kokugakuin.ac.jp

1. はじめに

本発表では、三重県鈴鹿市方言における 3 拍の名詞、動詞、形容詞のアクセントの実態を式の対立、所属語彙と型の相の観点から述べる。

三重県鈴鹿市は、服部（1930）の「近畿アクセントと東方アクセントとの境界線」や丹羽（2000）に述べられているように伝統的には京阪式のアクセントの地域である。しかし、近年報告があるように鈴鹿市では京阪式の体系を保持する話者と東京と似た体系を持つ話者が混在し、同一世代での個人差も大きい。このような複雑な当該地域のアクセントの実態を明らかにするために、拍数が比較的長く安定性があり、また古い文献による記録が残されており先行研究との比較も行いやすい 3 拍語に着目する。

三重県鈴鹿市のアクセントは、これまでに竹内(2015)(2017)で述べたように高起式、低起式という式の対立でみると以下の三つのタイプに解釈することができる。

(1) 式の対立と下がり目の位置で解釈できる A タイプ
(2) 拍数、型によって解釈の仕方が異なる AB タイプ（式の対立が曖昧）
(3) 下がり目の位置のみで解釈できるニュータイプ

本稿では上記の三つのタイプにわけ、三重県鈴鹿市における 3 拍語の実態を述べる。

A タイプ、AB タイプは式の対立の他にも異なる特徴が観察される。A タイプはそれぞれの型における所属語彙もほぼ京都と同様に観察される。それに対し、AB タイプは、共通語アクセント化が進んでいる様相が観察される。特に AB タイプにおいて共通語アクセントで話される語がどれだけあるかについては個人差がみられる。

2. 調査について

調査日は 2008 年 9 月から 2016 年 8 月で、調査は対面式で行った。動詞と形容詞は共通語での文例を示し、それを普段の言い方でどう言うかを尋ねる方法で行った。

話者は、15〜86 歳の鈴鹿市生え抜きの話者 42 名で、特に本稿では、A タイプの話者として H.O（1931 年生まれ、女性）を、比較的古相に近い AB タイプ①の話者として K.I（1985 年生まれ、女性)、比較的共通語化が進んでいる AB タイプ②の話者として S.T（1988 年生まれ、男性）の結果を中心に示す。

調査語は 1 拍から 4 拍の名詞と 2 拍から 4 拍の動詞と形容詞の活用形で、本稿では 3 拍の名詞 125 語、動詞 75 語、形容詞 25 語について詳述する。調査語彙は平山（1960）による。

調査地点である三重県鈴鹿市の人口は 20 万 282 人(2016 年 6 月 30 日)で、県北部に位置する。また、古くから東海道の要所として上方と江戸の行き来が盛んで、市内では 1616 年 44

番目の石薬師宿、1624 年に 45 番目の庄野宿が設けられた。

現在の交通としては、公共の乗り物は近畿日本鉄道(近鉄)が主で、乗り換えを考えると大阪、京都へ出るよりは名古屋方面へ向かう方が容易である。

以下記号については●、▶は高く発音される拍を、○、▷は低く発音される拍を、◎は中間的な高さを示す。以下の京都アクセント、東京アクセントは平山（1960）によるものとする。

3. 3 拍名詞のアクセント

3 拍名詞のアクセントについて、以下に式の対立、安定している型、タイプによって大きく異なるアクセントが観察される型についてそれぞれわけて詳述する。

3.1. 式の対立について

式の対立については、A タイプは対立を保持している相、AB タイプでは曖昧になっている相が観察される。具体的には表 1 のＡＢタイプの①にみられるように「鰻が」など伝統的には低起式の○○○▶となる語を安定して●●●▶で発音するなど伝統的な式とは異なる式で発音する語がみられる、「鼠が」「狐が」など●●●▶、○○○▶のどちらで発音してもよい語が観察されることがあげられる。

京都アクセントが高起式平板型●●●▶になる語は、ＡＢタイプにおいては一拍目が若干ゆるんで観察されるものの、低起式平板型より安定して京都と同じアクセントで観察される。ただし、ＡＢタイプ②の話者では「桜が」や「机が」が●●●▶、○○○▶のどちらで発音してもよいなど低起式平板型の「鼠が」「狐が」などと同様の様相も観察された。

表 1　京都アクセントが低起式平板型の語のアクセント

京都アクセント	類	語彙	Ａタイプ	ＡＢタイプ①	ＡＢタイプ②	東京式アクセント
○○○▶	1	昔	●●●▶	◎●●▶～●●●▶	◎●●▶～●●●▶	○●●▶
	6	兎	○○○▶	○○○▶	○○○▶	
		鰻	○○○▶	◎●●▶～●●●▶	○○○▶	
		狐	○○○▶	○○○▶～◎●●▶	○○○▶	
		スズメ	○○○▶	○○○▶	○○○▶	
		背中	○●○▷	○○○▶	○○○▶～○●○▷	
		鼠	○○○▶	○○○▶～◎●●▶	○○○▶	
		裸	○○○▶	○○○▶～◎●●▶	○○○▶	
		ヒバリ	○○○▶	◎●●▶～●●●▶	◎●●▶～●●●▶	
		誠	○○○▶	ＮＲ	○○○▶	
		ミミズ	○○○▶	○○○▶	○○○▶	

3.2　安定して観察される型について

京都アクセントと東京式アクセントが同じ語、たとえばどちらも頭高型●○○▷である語「えくぼが（2 類）」「鮑が（3 類）「鰈が（5 類）」などは全タイプを通じほぼ安定して頭高型で観察される。ただし、ＡＢタイプにおいては●○○▷～○●○▷でゆれが観察される語「枕が」「涙が」などが少数みられる。「枕が」「涙が」などは名古屋アクセントが○●

○▷となる語である。

京都アクセントが○●○▷である語については表 2 にあるように東京アクセントがどのような型であっても比較的安定して全年層を通じ○●○▷型を保持している。このような安定性は当該地域においては他の型にみられないもので、当該地域において○●○▷型は安定性があると考えられる。

表 2　京都アクセントが○●○▷型の語のアクセント

京都アクセント	類	語彙	Aタイプ	ABタイプ①	ABタイプ②	東京式アクセント
○●○▷	2	翼	●●●▶	◎●●▶～●●●▶	◎●●▶～●●●▶	○●●▶
		つるべ	ＮＲ	ＮＲ	◎●●▶～●●●▶	
		トカゲ	○●○▷	○●○▷	○●○▷	
		ムカデ	○●○▷	○●○▷	○●○▷	
	7	後ろ	○●○▷	○●○▷	○●○▷	
		鯨	○●○▷	○●○▷	○●○▷	
		薬	○●○▷	○●○▷	○●○▷	
		タライ	○●○▷	○○○▶	◎●●▶～●●●▶	
	2	緑	○●○▷	○●○▷	○●○▷	●○○▷
	6	高さ	○●○▷	○●○▷	○●○▷	
		狸	○●○▷	○●○▷	○●○▷	
	7	蚕	○●○▷	○●○▷	○●○▷	
		兜	○●○▷	○●○▷	○●○▷	
		便り	○●○▷	●○○▷	○●○▷	
		椿	○●○▷	○●○▷	○●○▷	
		病	○●○▷	●○○▷	○●○▷	
	2	毛抜き	○●○▷	○●○▷,●●●▷	○●○▷	○●●▷
		二つ	●●○▷～○●○▷	○●○▷	○●○▷	
		二人	●●○▷～○●○▷	○●○▷～●●○▷	○●○▷	

3.3　タイプによって大きく異なるアクセントが観察される型について

東京で○●●▷となる語は、当該地域のＡタイプとＡＢタイプで異なるアクセントが多く観察される。特にＡＢタイプでは京都アクセントでは観察されない尾高型◎●●▷～●●●▷が観察される。この尾高型は、初めの高さに関しては東京とは異なり語中であっても高く始まっても、若干ゆるんだ高さになっていてもよい。

また、同じＡＢタイプでも話者によって○●○▷と●○○▷で異なったアクセントで観察されていたり、個人内でも○●○▷～●○○▷でゆれていたりする語がある。これは楳垣（1957）、村中（2005）にあるように京阪式アクセントにおける●●○からの○●○や●○○への変化が関係していると考えられる。今回の調査において●●○▷型が観察されたのは、「小豆」「二つ」「二人」の 3 語であった。ただし、●●○▷のみが安定して観察されることはなく、3 語とも○●○▷でも発音される。しかし●●○▷型は 1985 年生まれのＡＢタイプでも観察された。どの語がどちらの型で発音される場合が多いのかなどは今後の課題である。

表 3　東京式アクセントで尾高型となる語のアクセント

京都アクセント	類	語彙	Aタイプ	ＡＢタイプ①	ＡＢタイプ②	東京式アクセント
●●●▶	1	麓	●●●▶	○●○▷～●●●▷～◎●●▷	○●○▷～●●●▷～◎●●▷	○●●▷
	4	恨み	●○○▷	●○○▷	●○○▷	
		暦	●○○▷	◎●●▶～●●●▶	○●○▷	
		林	●●●▶	◎●●▶～●●●▶	◎●●▶～●●●▶	
		仏	●●●▶	◎●●▶～●●●▶	◎●●▶～●●●▶	
●○○▷	2	小豆	●●○▷	●○○▷	●○○▷	
	3	力	●○○▷	●○○▷	●○○▷	
	4	頭	●○○▷	○●○▷	○●○▷	
		イタチ	○●○▷	○●○▷	◎●●▶～●●●▶	
		扇	●○○▷	○○○▶	○○○▶	
		男	●○○▷	●○○▷～○●○▷	●○○▷	
		面	●○○▷	ＮＲ	●●●▷～◎●●▷	
		鏡	●○○▷	○●○▷	○●○▷	
		仇	●○○▷	●○○▷～○●○▷	○●○▷	
		刀	●○○▷	○●○▷	○●○▷	
		瓦	●○○▷	●○○▷	○●○▷	
		宝	●○○▷	●●●▷～◎●●▷	●●●▷～◎●●▷	
		袴	●○○▷	○●○▷	○●○▷	
		東	●○○▷	◎●●▶～●●●▶	◎●●▶～●●●▶	
		光	●○○▷	●●●▷～◎●●▷	●●●▷～◎●●▷	
		響き	●○○▷	●●●▷～◎●●▷～●○○▷	●●●▷～◎●●▷	
		袋	●○○▷	○●○▷	○●○▷	
		筵(ムシロ)	●○○▷	ＮＲ	◎●●▶～●●●▶	
	5	襷	●○○▷	●○○▷	●○○▷	
○○○▶	2	昨夜(ゆうべ)	○○●▷	○○●▷	○○●▷	
○●○▷	2	毛抜き	○●○▷	○●○▷～●●●▷	○●○▷	
		二つ	●●○▷～○●○▷	○●○▷	○●○▷	
		二人	●●○▷～○●○▷	○●○▷～●●○▷	○●○▷	

4. 3 拍動詞・終止形のアクセント

鈴鹿市における 3 拍動詞のアクセントは、A タイプ、ＡＢタイプともに四つの型が観察された。●●●：高起式で下がり目のない型、○○●：低起式で下がり目のない型、○●○：低起式で 2 拍目に下がり目がある型、●○○：頭高型である。

●●●で観察されるのは主に 1 類の語、2 類の五段活用の語である。○○●で観察されるのは主に 3 類の語である。○●○で観察されるのは主に 2 類の一段活用の語である。京都では 2 類の一段活用の語は○○●で観察されるが、丹羽(2000)にあるように三重県鈴鹿市においては伝統的に○●○型で観察される。

表 4 にあるようにＡタイプとＡＢタイプで大きく異なるのは、Ａタイプにおいて○○●型で観察される語、すなわち主に 3 類の語である。具体的には「歩く（3 類)」「入る（3 類)」「遊ぶ（1 類)」がＡＢタイプでは○●○で発音されている。特にニュータイプでは、○○●型が観察されず、すべて○●○に変化している。

表 4　京都アクセントが○○●の動詞アクセント

京都アクセント	類	語例	Aタイプ	ＡＢタイプ①	ＡＢタイプ②	ニュータイプ	東京式アクセント
○○●	1A	遊ぶ	○○●	○○●	○○●	○●○	○●●
	1B	植える	○○●	○●○	○●○	○●○	
		捨てる	●●●	●●●～◎●●	●●●～◎●●	○●○	
	2B	生きる	○●○	○●○	○●○	○●○	○●○
		起きる	○●○	○●○	○●○	○●○	
		落ちる	○●○	○●○	○●○	○●○	
		掛ける	○●○	○●○	○●○	○●○	
		覚める	○●○	○●○	○●○	○●○	
		建てる	○●○	○●○	○●○	○●○	
		付ける	○●○	○●○	○●○	○●○	
		溶ける	○●○	○●○	○●○	○●○	
		撫でる	○●○	○●○	○●○	○●○	
		逃げる	○●○	○●○	○●○	○●○	
		晴れる	○●○	○●○	○●○	○●○	
	3	歩く	○○●	○○●	○●○	○●○	
		隠す	○○●	○○●	○●○	○●○	
	3	入る	○○●	○●○～○○●	○○●	○●○	

5.　3 拍形容詞・終止形のアクセント

表 5　3 拍形容詞のアクセント

京都アクセント	類	語例	Aタイプ	ABタイプ①	ABタイプ②	ニュータイプ	東京式アクセント	名古屋アクセント
●○○	1	赤い	●○○	●○○	○●○	○●○	○●●	○●○
		浅い	●○○	●○○	●○○	○●○		
		厚い	●○○	●○○	○●○	○●○		
		甘い	●○○	●○○	○●○	○●○		
		荒い	●○○	●○○	●○○	○●○		
		薄い	●○○	●○○	●○○	○●○		
		遅い	●○○	●○○	●○○	○●○		
		重い	●○○	●○○	●○○	○●○		
		堅い	●○○	●○○	●○○	○●○		
		軽い	●○○	●○○	●○○,○●○	○●○		
		暗い	●○○	●○○	●○○,○●○	○●○		
		遠い	●○○	●○○	●○○	○●○		
	2	熱い	●○○	●○○	●○○	○●○	○●○	
		黒い	●○○	●○○	○●○	○●○		
		白い	●○○	●○○	○●○	○●○		
		高い	●○○	●○○	●○○,○●○	○●○		
		近い	●○○	●○○	○●○	○●○		
		強い	●○○	●○○	○●○	○●○		
		長い	●○○	●○○	○●○	○●○		
		早い	●○○	●○○	○●○	○●○		
		低い	●○○	●○○	○●○	○●○		
		深い	●○○	●○○	○●○	○●○		
		古い	●○○	●○○	○●○	○●○		
		弱い	●○○	●○○	○●○	○●○		
	2	多い	●○○	●○○	●○○	○●○	●○○	

3 拍形容詞の京都アクセントは 1 類、2 類ともに●○○とされている。

鈴鹿市においても、Ａタイプ、すなわち高年層や古相を保つ若い世代は「赤い」「白い」など 1 類、2 類ともに●○○で観察される。ただし、ＡＢタイプでも共通語アクセント化が進んでいる話者やニュータイプにおいては 1 類、2 類ともに○●○型で発音する語が観察される。特に、式の対立を保持しておらず、異なる拍数や品詞で共通語アクセント化が進んでいるニュータイプの若年層話者においては 1 類、2 類ともに○●○のみが観察された。

芥子川(1983)にあるように、名古屋では東京のアクセントと異なり 3 拍形容詞は 1 類、2 類ともに○●○で発音される。みかけでは 3 拍形容詞・終止形のアクセントは名古屋アクセントと同じに変化している。ただし、名古屋アクセントの影響を受けたのか、頭高型のアクセントからアクセント核を一つ後ろに送る型への変化が進んだのかについては活用形のアクセントなどもふまえ今後の考察課題としたい。

6. おわりに

名詞は、式の対立が曖昧になりつつある様相が観察される。具体的には伝統的には低起式の○○○▶となる語を安定して●●●▶で発音するなど伝統的な式とは異なる式で発音する語がみられる、●●●▶、○○○▶のどちらで発音してもよい語が観察されるなどである。動詞は各年層を通じて古相を保持しており、類と型の対応がある。「歌う」や「余る」など 1 類と 2 類の五段活用の語が●●●、「生きる」など 2 類の一段活用の語が○●○、「歩く」など 3 類が○○●で対立を示し、式の対立を保持しているが、3 類の○○●は○●○に変化しつつある。形容詞は、高年層や古相を保つ若い世代は「赤い」「白い」など 1 類、2 類ともに●○○で観察される。ただし、変化が進んでいる新しいタイプの若年層は 1 類、2 類ともに○●○で観察される。品詞全体を通し○●○型は安定した型であるといえる。

参考文献

楳垣実（1957）「大阪方言アクセント変化の傾向」『近畿方言双書　方言論文集』2

芥子川律治(1983)「愛知県の方言」『中部地方の方言』（「講座方言学」6）国書刊行会

竹内はるか（2015）「三重県鈴鹿市のアクセントの研究」『國學院大學大學院紀要－文学研究科-』第 46 輯

竹内はるか（2017）「三重県鈴鹿市における用言の活用形のアクセント」『國學院大學大學院紀要－文学研究科-』第 48 輯

西宮一民（1961）「方言の実態と共通語化の問題点　三重・奈良・和歌山」東条操監修『方言学講座　第三巻　西部方言』東京堂

服部四郎（1930）「近畿アクセントと東方アクセントとの境界線」『音声の研究』第 3 輯　岡書院

平山輝男・丹羽一彌ほか（2000）『日本のことばシリーズ 24　三重県のことば』　明治書院

平山輝男編　（1960）『全国アクセント辞典』第 22 刷　東京堂出版

村中淑子（2005）「大阪における 3 拍名詞のアクセント－東大阪市 100 人調査の結果より－」『姫路獨協大学外国語学部紀要』18　姫路獨協大学

A3

漢語類別語彙の検討

（Examination of the Classification of Chinese Characters on Accentuation）

秋山　英治（愛媛大学法文学部）
akiyama.eiji.mk@ehime-u.ac.jp

1. はじめに

日本語諸方言アクセントにおいて，金田一（1980）や奥村三雄氏の一連の研究によって，漢語にも和語に準じる類別語彙があることが知られている．しかし，この語彙について検証した研究は，管見の範囲では中井（1997b）がある程度である．そのため，漢語類別語彙が，諸方言間の対応を示すものとして使用できるものであるのか，よくわかっていない．本発表では，漢語類別語彙について現代方言をもとに比較検討した結果を述べる．

2. 漢語類別語彙候補語の抽出

漢語類別語彙については，金田一（1980）と奥村氏の一連の研究で，類のたて方，類の所属語などに違いがみられる．金田一（1980）は，1・2 拍語の記述しかなく，3 拍語はない．また，1 拍語では第 2 類，2 拍語では第 2・5 類を欠く．一方，奥村氏の一連の研究（1955・1961・1963・1964・1974・1981・1990）では，研究によって類語としてあがっている語に違いがあり，また同じ語でも研究によって類の認定が異なるものもある．さらに，奥村（1961）では，2 拍第 2・3 類が区別されず同じ類として記述されており，類の認定が難しいものもある．

そこで，金田一（1980）および奥村氏の一連の研究（1955・1961・1963・1964・1974・1981・1990）より，現代では使用しないと思われるものを除き，どれか 1 つでも類語としてあがっていれば，それを検討対象語（「候補語」）として抽出した．3 拍語については，金田一（1980）に記述がないことから，奥村氏の一連の研究より抽出した．類の認定は，奥村氏の一連の研究にしたがった．金田一（1980）と奥村氏の一連の研究とで類が異なる語は，奥村氏の類を採った．奥村氏の一連の研究のなかで，類が異なる語については，一番新しい研究の類を採った．奥村（1961）で，区別されていない 2 拍第 2・3 類については，その後の研究で類が特定された場合はその類を採用し，その後の研究に記述がない場合は暫定的に第 2 類とした．「鉢（ハチ）」など和語の類別語彙のなかに含まれている語もあるが，金田一（1980）および奥村氏の一連の研究であがっていれば，それも候補語として抽出した．なお，一部の語については，秋永一枝ほか（1998）も参照した．

以上より，抽出した候補語は，合計 865 語となる．拍数別にみると，1 拍語が 58 語（第 1 類 34 語，第 3 類 17 語，その他 7 語）で，第 2 類を欠く．2 拍語は 325 語（第 1 類 79 語，第 2 類 58 語，第 3 類 63 語，第 4 類 72 語，第 5 類 5 語，その他 48 語）で，すべての類に候補語がある．しかし，第 5 類は 5 語しかなく，他の類と比べて著しく少ない．3 拍語は 473 語（第 1 類 90 語，第 4 類 71 語，第 5 類 164 語，第 6 類 75 語，第 7 類 71 語，その他 2

語）で，第 2・3 類を欠く．和語においても，3 拍第 3 類を類としてたてるかどうか，研究者によって意見がわかれるが，それを除くと，和語は 1〜3 拍語すべてに類がたてられているのに対して，漢語は 1・3 拍語で類を欠いており，和語と同様に扱うことができないことがわかる．以下，紙面の都合上，2・3 拍語に限定して述べていく．

3. 対象地域

今回対象とした地域は，「中央式」から京都市・高知市・徳島市・今治市の計 4 地域，「讃岐式」から丸亀市（飯岡）の 1 地域，鹿児島，共通語である．これらの地域での状況を比較し，類語として妥当かどうかを検証する．「中央式」のうち，京都市のデータは，平山編（1960）と中井（2002）による（前者を京都①，後者を京都②とする）．高知市のデータは中井（1997a），徳島市のデータは中井（1999）による．今治市のデータは，筆者の調査（老年層 1 名．秋山（2017）の話者と同じ）による．「讃岐式」の丸亀市のデータは，中井（1998）による．鹿児島のデータは，平山編（1960）による．共通語のデータは，平山編（1960）と金田一監修（2014）による（前者を共通語①，後者を共通語②とする）．各地域の型を示すと，次のようになる．

「中央式」（京都市・高知市・徳島市・今治市）　※京都市では，第 4 類は H1 型となる
2 拍語　第 1 類 H0／第 2・3 類 H1／第 4 類 L0／第 5 類 L2
3 拍語　第 1 類 H0／第 4 類 H2※／第 5 類 H1／第 6 類 L0／第 7 類 L2
「讃岐式」（丸亀市）　※第 5 類は H1 型も
2 拍語　第 1・3 類 H0／第 2 類 H1／第 4 類 L0／第 5 類 L2
3 拍語　第 1・5 類 H0※／第 4 類 H1／第 6 類 L0／第 7 類 L2
鹿児島（平安末頃京都で高起始まりの語が A 型，低起始まりの語が B 型）
2 拍語　第 1・2 類 A／第 3・4・5 類 B
3 拍語　第 1 類 A／第 4・5・6・7 類 B
共通語　　第 5 類は，現在 1 型となっている語が多いが，本来の 2 型として考察する．
2 拍語　第 1 類 0／第 2・3 類 2／第 4・5 類 1
3 拍語　第 1・6 類 0／第 4 類 3／第 5 類 2※／第 7 類 1

4. 新・漢語類別語彙（暫定版）の認定

各地域の比較検討により，「その類の語として認定できる語（類語）」「類語に準じる語（準類語）」，「その類の語として認定できない語（非類語）」「今回の調査では類語かどうか確定できずさらなる調査が必要な語」にわけた．以下では，「類語」「準類語」を示す．また，金田一（1980）・奥村氏の研究であがっている類とは別の類の可能性がある語（分類上は，「非類語」に含む）がある場合は，それも示す．各地域で例外的な型をとる場合，その型のみであれば，（　）に地域・型を記し，併用に例外的な型があれば，型の後に「も」を付す．副詞的用法など用法により型が異なる可能性のある語は，※を付す．一部の地域のデータしかない語（全地域の比較ができないため，「準類語」と判断した語）は，■を付す．

2 拍語

第 1 類（候補語 79 語）

類語……医者(鹿児島 B)・椅子(高知・徳島 H1)・梅・客(今治 H1 も)・逆(鹿児島 B)・急・胡麻(鹿児島 B)・三※・邪魔(鹿児島 B)・棕櫚(共通語①1 も)・順・食(京都②・今治 H1 も)・宅(今治 H1)・敵(高知 H1)・点(高知 H1 も，徳島 H1)・得(今治 H1 も）・晩・無事(京都①L0，京都②L0)・棒　　計 19 語

準類語…格■(京都②・今治 H1)・曲■(丸亀 H1，共通語①・②1 も)・情(徳島・今治 H1，鹿児島 B)・職(京都②H1 も，東京①2 も)・席(共通語①1 も，共通語②1)・俗(今治 H1，共通語①1 も)・賊(京都①・今治 H1)・徳(高知・今治 H1)・櫃(徳島 L0，今治 H1 も，鹿児島 B)・表(京都②L0 も，今治・丸亀 H1)・塀(京都②・今治 H1 も，徳島 H1，鹿児島 B)・偏(京都・今治 H1 も，徳島 H1)・陸(共通語①・②2 も)・礼※(京都②L0 も，共通語②1 も)・碌(今治 H1，共通語②2 も)　　計 15 語

別の類の可能性がある語（6 語）

第 2 類の可能性がある語…興・骨・産　　第 3 類の可能性がある語…作

第 4 類の可能性がある語…灸・訓

第 2 類（候補語 58 語）

類語……意地・七・余所(共通語①1 も)　　計 3 語

準類語…御座(鹿児島 B，東京②1)　　計 1 語

別の類の可能性がある語（10 語）

第 3 類の可能性がある語…一※・画・義理・損・檀・知恵・八※・百※・風呂・六※

第 3 類（候補語 63 語）

類語……寒・菊・柵(共通語①・②1 も)・桟・軸・質(徳島 L0)・尺(丸亀 H1 も)・数珠(高知 L2 も)・性(京都②H0 も)・世話(丸亀 H1 も)・段・弟子・毒・肉・二度※・熱・鉢・罰・撥・便・服(丸亀 H1)・福・副(丸亀 H1)・幕・膜・脈(今治 H0 も)・約(共通語②0)・欲・楽(鹿児島？)・牢・蝋・枠　　計 32 語

準類語…餓鬼(今治・丸亀 L2)・吉（丸亀 H1，共通語②1 も）・柵(京都②H0 も，丸亀 H1 も，共通語①1，共通語②1 も)・式(丸亀 H1)・実(徳島 L2 も，丸亀 H1)・塾・術(丸亀 H1，共通語①1 も)・用(丸亀 H1)　　計 8 語

第 4 類（候補語 72 語）

類語……運・会・害・経・王(京都②・今治・丸亀 H1 も)・恩(今治 H1 も)・芸・剣(京都②・丸亀 H1 も)・十(今治 H1)・芯・千(今治 L2)・台・題・中(今治 H1)・腸(今治 H1 も)・胴・堂(京都②・徳島・今治 H1 も)・判(京都②H1 も，高知 H1)・番(今治 H1)・盤(今治 H1 も，丸亀 H1)・糞・味噌・紋・例(鹿児島 A)　　計 24 語

準類語…縁※(今治 H1)・癌(京都①・高知 H1，京都②・今治・丸亀 H1 も)・香(京都①・徳島・今治 H1，京都②H1 も)・杓(今治 H1 も，共通語①2 も)・精(高知・今治 H1，丸亀 H1 も，鹿児島 A)・代(今治 H1，東京②0)・宙(今治・丸亀 H1 も，東

京①・②0 も)・町(高知・丸亀 H1，今治 H1 も)・同(京都②・今治 H1 も，高知 H1，鹿児島？)・二個■・念(京都①・今治・H1，共通語①0 も)・盆(鹿児島 A，共通語①0)・瘍(京都①・高知・今治 H1)・類(京都①H1，京都②H1 も，鹿児島 A)・論(高知・徳島 H1，京都②・丸亀 H1 も)　計 15 語

別の類の可能性がある語（1 語）　第 3 類の可能性がある語…陣

第 5 類（候補語 5 語：雁・犀・皮膚・豹・法螺）　※類語・準類語なし

3 拍語

第 1 類（候補語 90 語）

類語……遠慮・機嫌・麒麟・工夫・故障(高知 H2)・事情(丸亀 H1)・時代(今治 L0 も)・修行(鹿児島 B も)・障子・女中・他人・達摩・調子・内緒・秘密(京都②L0 も)・不思議・葡萄(鹿児島 B)・魔法(丸亀 H1)・微塵・模様・油断(京都②L0 も，高知 L0)・用事(鹿児島 B)・陸地(丸亀 L0)・理屈　計 24 語

準類語…九年■・五年■(京都②L2 も)・五杯■・五番■・五枚■・詩人(京都②H1 も，高知 L0)・草履(丸亀 L0，鹿児島 B)・達者(今治 L0，鹿児島 B)・法事(京都②H2 も，徳島 H1)　計 9 語

別の類の可能性がある語（1 語）　第 6 類の可能性のある語…相違

第 4 類（候補語 71 語）

類語……愛想(徳島 H0 も)・道具・二月(共通語②0 も)　計 3 語

準類語…脚気(共通語①0)・四角(京都①L2，京都②L2 も)・四月(共通語②0 も)・地獄(京都②L0 も，共通語①0)・上手(東京①0)・豆腐(京都②L0 も，共通語①・②0 も)・道理(徳島 L2，丸亀 H1)・難儀(共通語②1 も)・二百※・拍子(鹿児島 A，共通語①・②0 も)　計 10 語

第 5 類（候補語 164 語）

類語……該当なし　計 0 語

準類語…五寸※(共通語②0 も)・五万※(鹿児島 A)・秩序(丸亀 H1，共通語①・②1 も)・不孝(京都②L2 も，丸亀 L2)　計 4 語

別の類の可能性がある語（8 語）

第 4 類の可能性がある語…苦心・砂糖・七分※・世界・二点※・日本・二万※・八位

第 6 類（候補語 75 語）

類語……按摩(高知 H0)・硫黄・縁起(高知 H0)・火災・漢字・桔梗(京都②H0 も)・最初・丈夫・真珠(高知 L2)・掃除・太鼓・檀那・茶碗・二階・二重・屏風・楊枝・蓮華(京都②H1 も)　計 18 語

準類語…増加(京都②H1・L2 も，共通語①1 も)・題字■・檀家■・二乗■・柔和■・飛脚■・陽気■・幼稚(共通語①・②1 も)　計 8 語

第 7 類（候補語 71 語）

類語……昆布(京都②L3 も)・今夜・最後(丸亀 H1)・坊主・名字　計 5 字

準類語…王子(オウジ)■・近所(キンジョ)(高知・丸亀 H0，京都②H0 も)・軍事(グンジ)■・三里(サンリ)※(高知 H0 も，徳島 L0 も)・千里(センリ)※(徳島 L0 も，丸亀 H1 も)・文句(モンク)(丸亀 H1，鹿児島 A)・用意(ヨウイ)(京都①L0，共通語①0)　計 7 語

2 拍第 2 類・3 拍第 5 類については，類語・準類語が少なく，独立した類としてたてることが難しい.これは，中井（1997b）が指摘するように，漢語の基本型と区別できないことによる．2・3 拍語ともに基本型が出やすい傾向がある上に，2 拍第 2 類は，候補語に 2 拍目に特殊音となる語が多く，共通語においては，特殊音により下げ核が前へずれたのか，もともとの型（1 型）なのかを判断することが難しい.3 拍第 5 類は，候補語として最多の 164 語があがっていたものの，類語として認定できる語はなく，準類語が 4 語あるだけである．

候補語の段階で 5 語しかなかった 2 拍第 5 類については，類語がなく類をたてることができない．その他，今回の調査では類ができない語が多く残るといった問題があり，暫定版ではあるものの，各地域の比較検討により，上記の新・漢語類別語彙の語を認定した．

5. 新・漢語類別語彙（暫定版）の検証

新・漢語類別語彙（暫定版）が，諸方言間の対応を示すものとして，また世代間の対応を示すものとして使用できるものかどうかを検証するために，愛媛県東中予地方でもっとも共通語化が進行している松山市方言の若年層（1 名）をとりあげ，和語との比較をおこなった．具体的には，漢語にも，和語と同様の変化（共通語化）が起きているのかを確認した．なお，ここでは，別の類の可能性がある語はとりあげず，類語・準類語をとりあげる．

2 拍語では（独立した類として認定の難しい第 2 類，類語のない第 5 類は考察対象から除く），共通語化により，和語の第 2・3 類に 2 型，第 4・5 類に 1 型がみられる．同様に漢語にも，第 3 類に 2 型，第 4 類に 1 型がみられ，和語とともに共通語化が起きている．

3 拍語では，第 7 類に，和語・漢語ともに 1 型がみられ，和語とともに共通語化が起きているが，第 4 類は，和語と漢語で振る舞いが異なる．和語の第 4 類では，26 語中 22 語に共通語型の 3 型がみられる（総語数に対する 3 型の出現率が 85%）のに対して，漢語第 4 類では，3 型が 14 語中わずか「愛想」の 1 語（出現率 7%）のみである．松山市では，語末核型が少なく，3 拍第 4 類（和語）の共通語化が，伝統的な H1 型から非共通語型の 2 型を挟んで共通語の 3 型へという独特の変化を起こしている（秋山 2017）ことから，共通語の 3 型に変化する前の中間段階として，単語単独で同じ型の 0 型をとったということも考えられなくはない．しかし，「愛想」を除く 13 語のうち，8 語が 0 型で，中間段階の型が和語と違う．また，漢語の 3 拍語に，第 4 類以外の語にも語末核型の 3 型が数語ほどしかみられない．2 拍語には多くみられる語末核型が 3 拍語にほとんどみられないのは不審である．漢語，とくに 3 拍第 4 類に語末核語がみられない原因については，複合語規則が関与している可能性が考えられる．現時点で断言はできないが，漢字 1 字で表す語が多い和語に対して，漢語では漢字 2 字あるいは 3 字で表すため，複合語としての意識が働いていた可能性がある．この点については，今後さらに詳細な調査をおこない，明らかにしたい．いずれにせよ，漢語の 3 拍第 4 類を，諸方言間の対応を示すものとして扱うことには問題がある．

6. おわりに

以上より，漢語類別語彙について，2 拍第 2・5 類，3 拍第 2・5 類を欠くこと，また 3 拍第 4 類が和語と異なる振る舞いをみせていることなどから，諸方言間（世代間）の対応を示すものとして使用するには問題があることが明らかになった．今後は，今回認定した語について，他地域・他世代のデータを広く収集し，再確認するとともに，今回データ不足から認定できなかった語についてもとりあげ，漢語類別語彙の妥当性をさらに検証したい．

附記

本研究は，JSPS 科研費 17K02733 の助成を受けたものである．

参考文献

秋永一枝・上野和昭・坂本清恵・佐藤栄作・鈴木豊編(1998)『日本語アクセント史総合資料 研究篇』東京：東京堂出版

秋山英治(2017)『愛媛県東中予方言のアクセントと共通語のアクセント-日本語史再建のために-』東京：おうふう

奥村三雄(1955)「東西アクセント分離の時期-外来語のアクセント-」『国語国文』24:12,34-44.

奥村三雄(1961)「漢語のアクセント」『国語国文 30:1,1-16.

奥村三雄(1963)「漢語のアクセント-アクセントから語彙論へ-」『国語学』55,36-53.

奥村三雄(1964)「漢語アクセントの一性格」『国語国文』33:2,48-68.

奥村三雄(1974)「諸方言アクセント分派の時期-漢語アクセントの研究-」『方言学研究叢書 3』,1-38.

奥村三雄(1981)『平曲譜本の研究』東京：桜楓社

奥村三雄(1990)『方言国語史研究』東京：東京堂出版

金田一春彦(1980)「味噌よりは新しく茶よりは古い-アクセントから見た日本祖語と字音語-」『言語』9:4,88-98.

金田一春彦監修・秋永一枝編(2014)『新明解日本語アクセント辞典 第 2 版 CD 付き』東京：三省堂

中井幸比古(1997a)『高知市方言アクセント小辞典-方言アクセント小辞典(1)-』（科研報告書）（中井幸比古(2002)に再録）

中井幸比古(1997b)「讃岐式アクセントと中央式アクセントの対応について」『島田治還暦記念論文集－言葉と文化－』,89-111.（中井幸比古(2002)に再録）

中井幸比古(1998)『香川県方言アクセント小辞典-方言アクセント小辞典(2)-』（科研報告書）（中井幸比古(2002)に再録）

中井幸比古(1999)『徳島市方言アクセント小辞典-方言アクセント小辞典(3)-』（科研報告書）（中井幸比古(2002)に再録）

中井幸比古(2002)『京阪系アクセント辞典』東京：勉誠出版

平山輝男編(1960)『全国アクセント辞典』東京：東京堂出版

A4

The phonetic reduction of nasals and voiced stops in Japanese across speech styles

Yoichi Mukai, Benjamin V. Tucker (University of Alberta)
mukai@ualberta.ca, bvtucker@ualberta.ca

1 Introduction

In daily conversations, speakers often produce speech in a casual manner. Casual speech, also referred to as spontaneous or conversational speech, contains a high degree of variation as compared to more careful speech styles (Ernestus and Warner, 2011). One important aspect of casual speech leading to this high variability is phonetic reduction, resulting in words being pronounced with fewer segments, shorter durations, and assimilation. For example, *yesterday* pronounced carefully could be something like /jɛstə˞deɪ/ but in casual speech it could be pronounced [jɛʃeɪ] (Tucker, 2007). Reduced pronunciation variants has been studied cross-linguistically with evidence being reported in American English (e.g., Johnson, 2004; Warner and Tucker, 2011), Dutch (e.g., Ernestus et al., 2002), French (e.g., Brand and Ernestus, 2015), Finnish (e.g., Lennes et al., 2001), German (e.g., Kohler, 1990), and Japanese (e.g., Arai et al., 2007; Maekawa, 2005).

In the present study, we use a large-scale speech corpus, the Corpus of Spontaneous Japanese (Maekawa, 2003), to examine the phonetic variability found in nasals and voiced stops and to describe how that variation and reduction occurs across speech styles in Japanese. Using the Corpus of Spontaneous Japanese, we analyzed the duration and intensity difference of target segments across four styles of speech: academic presentations, simulated public speech, dialogues, and read speech. The intensity difference was defined as the difference between the minimum intensity of the target segment to the averaged maximum intensity of surrounding segments (Tucker, 2011; Warner and Tucker, 2011). We hypothesized that we would observe stronger reduction (more approximant-like productions), as indicated by shorter duration and smaller intensity difference, as speech style becomes more casual. In other words, the shortest duration and the smallest intensity difference would be found for nasals and voiced stops in dialogues (most casual) and the longest duration and the largest intensity difference in read speech (least casual).

2 Methods

2.0.1 Data

We used the Corpus of Spontaneous Japanese, which contains approximately 44 hours of speech (about half million words) from four different speech styles: academic presentations, simulated public speech, dialogues, and read speech (Maekawa, 2003). All acoustic analysis was performed using Praat (Boersma and Weenink, 2016) and the predefined segmental boundaries provided in the corpus.

2.0.2 Analysis

We used linear mixed-effects models with lme4 and lmerTest packages (Bates et al., 2017; Kuznetsova et al., 2015) in R (R Core Team, 2017) to measure whether duration and intensity difference of nasals and voiced stops differs across speech styles, as well as to predict the relative duration and intensity difference of these segments across speech styles. Furthermore, we also ran Bonferroni adjusted post hoc comparisons between SpeechStyle and Phoneme using the multcomp and lsmeans packages in R (Lenth, 2017; Torsten Hothorn, 2016). The variables of interest were as follows:

- Dependent variables: LogDuration; LogIntensityDifference (A log-transformation was applied to attenuate skewness)
- Main predictors: Phoneme (nasals: /m/, /n/, /ɴ/; voiced stops /b/, /d/, /g/); SpeechStyle (AcademicPresentation, SimulatedPublicSpeech, Dialogue, and ReadSpeech)
- Control variables: SpeakerAge, WordDuration, and PhonemeEnvironment (Word-initial, -medial, -final)

All the control variables were included in the models as long as the variables significantly contributed to the fit of the model. We also included Speaker as a random intercept and SpeechStyle by Speaker as a random slope.

3 Results & Discussion

3.1 Duration

Statistical analysis of nasal duration across speech styles, visualized in Figure 1, revealed that there is a main effect of SpeechStyle [$F(3,8)=13.3$, $p<0.001$] and Phoneme [$F(2,138229)=2829.1$, $p<0.001$] as well as an interaction between SpeechStyle and Phoneme [$F(6,137122)=44.0$, $p<0.001$]. Individual comparisons of /m/ across speech styles revealed that the duration of /m/ is longer for the read speech in comparison to simulated public speech (t=-4.7, $p<0.001$) and academic presentations (t=-5.97, $p<0.001$). The /m/ nasal durations in simulated public speech are also significantly longer than in academic presentations (t=3.77, $p<0.01$). The comparisons between dialogues and academic presentations, as well as dialogues and simulated public speech, are not significant. We also identified that /ɴ/ follows a similar pattern to /m/ where the

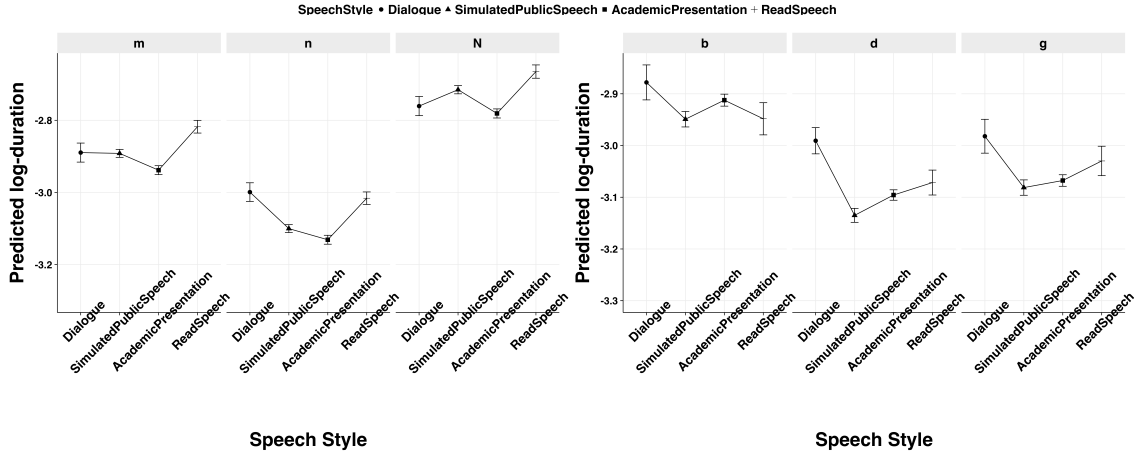


Figure 1: Interaction effect of Phoneme and SpeechStyle on log-duration

duration of /ɴ/ for read speech is the longest among the other speech styles [Dialogue and ReadSpeech (t=-3.61, $p<0.01$); AcademicPresentation and ReadSpeech (t=-5.56, $p<0.001$); SimulatedPublicSpeech and ReadSpeech (t=-3.05, $p=0.05$)]. We also found that the /ɴ/ durations of simulated public speech are significantly longer than academic presentations (t=5.24, $p<0.001$). The comparisons between dialogues and academic presentations, as well as dialogues and simulated public speech, are not significant. Analyses of /n/ revealed a different pattern from both /m/ and /ɴ/ in which the /n/ duration for both dialogues and read speech are longer (no significant difference between Dialogue and ReadSpeech) than simulated public speech and academic presentations [Dialogue and SimulatedPublicSpeech (t=4.97, $p<0.001$); Dialogue and AcademicPresentation (t=4.55, $p<0.001$); SimulatedPublicSpeech and ReadSpeech (t=-5.52, $p<0.001$); AcademicPresentation and ReadSpeech (t=-5.77, $p<0.001$)]. The comparisons between dialogues and read speech, as well as academic presentations and simulated public speech, are not significant. Figure 1 indicates in the left panel that dialogues show the shortest durations and read speech displays the longest durations for both /m/ and /ɴ/, as predicted. However, simulated public speech and academic presentations indicate no difference or an opposite relationship and does not fit with our prediction, in which simulated public speech should have longer duration than academic presentations. A possible explanation for this is that academic presentations are often prepared speech that are highly contentful, allowing speakers to deliver a very dense talk faster than unprepared speech. This might be the source of the shorter durations in academic presentations. Research has also found that the faster speech rate leads to more instances of reduction (Brand and Ernestus, 2015). Additionally, as seen on the left side of Figure 1 in the middle panel, /n/ shows an unexpected pattern in which there is no difference between dialogues and read speech, and both simulated public speech and academic presentations display shorter duration than dialogues and read

speech. It is possible that there is a particular word that is common in dialogues but that does not occur in the other speech styles causing this effect. However, further analysis is needed to explore this pattern and confirm our hypotheses about this effect.

Analyses of the duration of voiced stops across speech styles also revealed a main effect of SpeechStyle [F(3,10)=7.53, $p<0.01$] and Phoneme [F(2,37910)=125.99, $p<0.001$] as well as an interaction between SpeechStyle and Phoneme [F(6,36380)=3.40, $p<0.01$]. Voiced stops display different patterns, both from what we predicted and the pattern of nasals. Individual comparisons of the /b/ duration across speech styles revealed that all the comparisons are not significant. We also found that the duration of /d/ is longer for dialogues than for both academic presentations (t=3.57, $p<0.05$) and simulated public speech (t=6.05, $p<0.001$), but other comparisons are not significant. Although /g/ shows a similar pattern to /d/, none of the comparisons among the speech styles in /g/ are significant in this data set. As shown on the right side of Figure 1, the results do not follow our predictions for speech style. In addition, fewer comparisons in voiced stops reached significance as compared to in nasals. In order to further examine these tendencies, we conducted additional analyses by segmenting the stops into two parts: closure duration and release duration. The closure duration was defined as the duration from the offset of the preceding segment to the onset of a burst release, and the release duration was defined as the duration from the onset of a burst release to the onset of the following segment. However, due to the approximated articulation of stops, especially for dialogues, a number of the stops did not have closure and release durations as a separate unit, meaning that the boundary between the offset of closure duration and the onset of release duration was unclear. Table 1 illustrates the numbers of the stops with clear boundary between closure and release durations and with no boundary between the two durations across speech styles. As expected, the number of the stops with no boundary is the highest in dialogues and the lowest in read speech. For the following analysis, the stops with clear boundary (83,957 stops) were utilized.

Table 1: Number of voiced stops with ClearBoundary and NoBoundary and their ratio

SpeechStyle	ClearBoundary	NoBoundary	Total	BoundaryRatio	NoBoundaryRatio
Dialogue	4164	236	4400	94.64	5.36
SimulatedPublicSpeech	39839	1889	41728	95.47	4.53
AcademicPresentation	36556	1408	37964	96.29	3.71
ReadSpeech	3398	40	3438	98.84	1.16

3.1.1 Closure duration

Figure 2 illustrates an interaction between phonemes and speech style on release and closure durations. Statistical analysis of closure duration across speech styles demonstrated a main effect of SpeechStyle [F(3,15)=3.85, $p<0.05$] and Phoneme [F(2,36253)=184.78, $p<0.001$] as well as an interaction between SpeechStyle and Phoneme [F(6,33554)=5.936, $p<0.001$]. Individual comparisons of /b/ across speech styles revealed that none of the differences in closure duration across speech styles reached significance. For /d/, we found that relationships among the speech styles differ from that of /b/ where the closure duration of /d/ for dialogues is longer than for simulated public speech (t=7.07, $p<0.001$), and we also identified that the /d/ closure duration for simulated public speech is shorter than for both academic presentations (t=-4.0500, $p<0.01$) and read speech (t=-3.82, $p<0.01$). The other comparisons are not significant. The /g/ closure duration displays similar relationships among the speech styles to /b/ but the degree of durational differences across speech style are small; therefore, none of the differences reached significance. Figure 2 in the left panel indicates that closure duration shows a similar pattern to the entire stop duration where the durations are the longest in dialogues and the shortest in simulated public speech, but the degree of durational differences across speech styles for closure durations are smaller than for the entire stop durations.

3.1.2 Release duration

Likewise, analyses of release duration across speech styles revealed that there is an main effect of SpeechStyle [F(3,12)=4.12, $p<0.05$] and Phoneme [F(2,38921)=578.30, $p<0.001$] as well as an interaction effect between SpeechStyle and Phoneme [F(6,38632)=15.16, $p<0.001$]. We identified that the release duration of /b/ in dialogues is the longest among the other speech styles [Dialogue and SimulatedPublicSpeech (t=4.12, $p<0.01$); Dialogue and AcademicPresentation (t=4.83, $p<0.001$); Dialogue and Read-

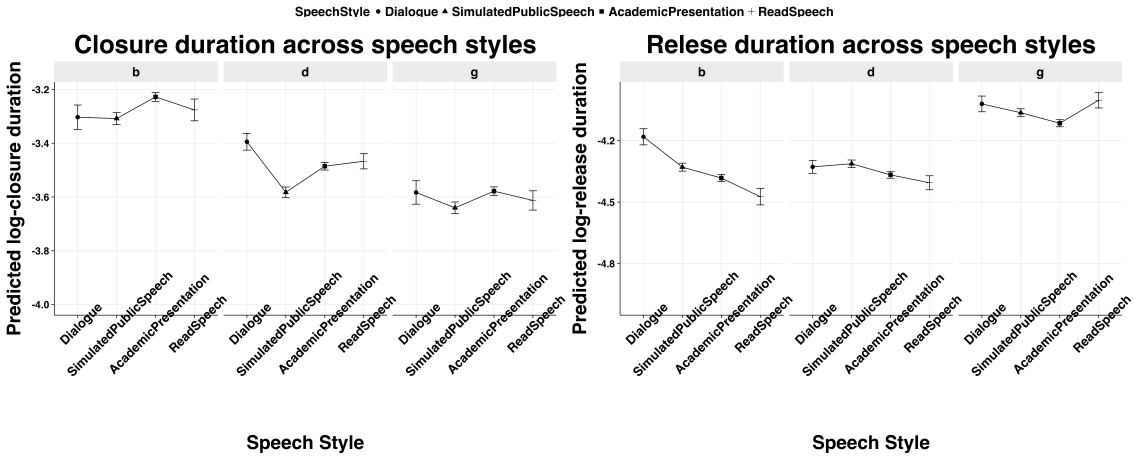


Figure 2: Interaction effect of Phoneme (voiced stops) and SpeechStyle on log-release and closure duration

Speech (t=5.17, p<0.001)]. We also found that the /b/ release duration for simulated public speech is longer than for read speech (t=4.06, p<0.01). The comparisons between academic presentations and read speech, as well as academic presentations and simulated public speech, are not significant. Individual comparisons of /d/ across speech styles revealed that release duration of /d/ displays similar relationships among the speech styles to /b/ but the degree of the differences are smaller. The /d/ release duration for simulated public speech is longer than for read speech (t=3.30, p<0.05), but all other comparisons did not reach significance. As shown by Figure 2 in the right panel, the /g/ release duration also indicates similar relationships among the speech styles to the other two stops except that read speech is relatively longer. Similarly, the degree of durational differences for /g/ across speech styles are small; therefore, none of the differences reached significance. As in the results of the entire duration of voiced stops, we identified an overall tendency towards the release duration being longer in casual speech (i.e., dialogues and simulated public speech) as compared to careful speech (i.e., academic presentation and read speech) except /g/. Interestingly, although the overall tendency was opposite to what we expected, release duration displayed consistent durational differences across speech styles. Additionally, the number of significant differences were greater than that of the entire duration of voiced stops. As a result, our findings here suggest that speech style difference is better reflected in release duration than the entire duration of the voiced stop. Importantly, as shown by the both panels of Figure 2, the relationships among the speech styles in closure and release durations vary, suggesting that the way in which speech styles impact release and closure durations differ.

3.2 Intensity difference

Figure 3 illustrates an interaction between phoneme and speech style on intensity difference. Analyses of intensity difference in nasals across speech styles demonstrated that there is a main effect of SpeechStyle [F(3,12)=16.8, p<0.001] and Phoneme [F(2,138612)=5168.7, p<0.001] as well as an interaction between SpeechStyle and Phoneme [F(6,137983)=51.7, p<0.001]. Individual comparisons of /m/ across speech styles revealed that the intensity difference of /m/ in academic presentations are greater than in read speech (t=3.73, p<0.01). Other comparisons did not reach significance. Analysis of /n/ revealed that relationships among the speech styles in /n/ is similar to that of /m/ in which the /n/ intensity difference for dialogues is smaller than for academic presentations (t=-3.24, p<0.05). We also found that the /n/ intensity difference for simulated public speech is greater than both for read speech (t=3.43, p<0.05) and academic presentations (t=7.42, p<0.001). The /ɴ/ intensity difference shows different relationships among the speech styles from that of both /m/ and /n/ but none of the differences reached significance. As shown by Figure 3 in the left panel, /m/ and /n/ display a similar pattern in which dialogues contain a smaller intensity difference as compared to both simulated public speech and academic presentations, as well as simulated public speech possesses a smaller intensity difference as compared to academic presentations. However, unlike our prediction, the intensity difference of read speech is smaller than that of both simulated public speech and

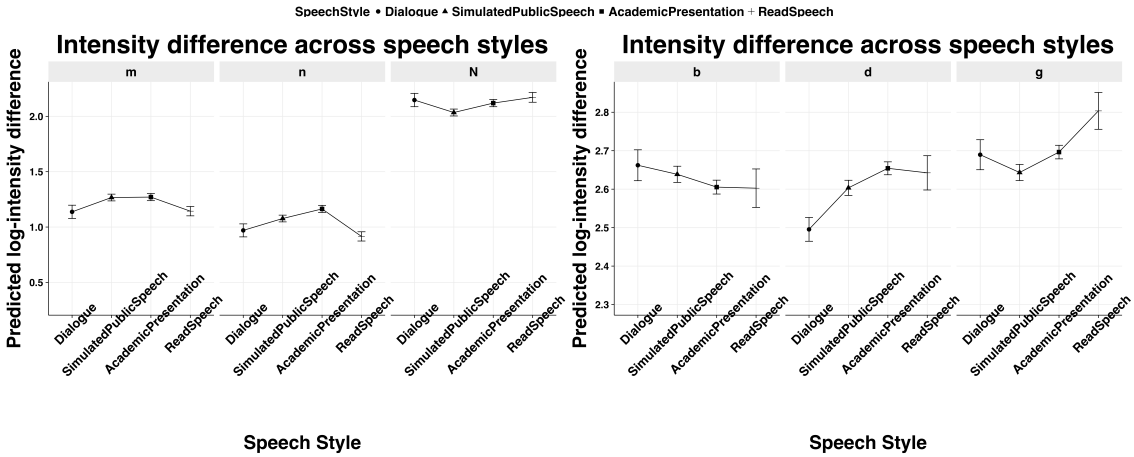


Figure 3: Interaction effect of Phoneme and SpeechStyle on log-intensity difference

academic presentations in both /m/ and /n/. Additionally, the intensity difference of /ɴ/ are similar across speech styles except that simulated public speech shows a smaller intensity difference than the others. A possible explanation for this could be that due to the sonority hierarchies, the intensity difference between stops (oral stops) and neighbouring sounds (most likely vowels) is expected to be smaller than the intensity difference between nasals (nasal stop) and neighbouring sounds (most likely vowels). This means that inherently voiced stops have more space to indicate reduction by a smaller intensity difference than nasals. It is possible that the intensity difference measurement is more appropriate to measure reduction for voiced stops (oral stops) than nasals (nasals). This account is reflected in the results of voiced stops.

Analyses of intensity difference in voiced stops revealed that there is no main effect of SpeechStyle but there is a main effect of Phoneme [$F(2,38816)=61.61$, $p<0.001$] and an interaction between SpeechStyle and Phoneme [$F(6,38831)=25.55$, $p<0.001$]. Individual comparisons of /b/ across speech styles indicate that none of the differences among the speech styles reached significance. Analyses of the /d/ intensity difference show that the relationships among the speech styles in /d/ differ from that of /b/ where the /d/ intensity difference for dialogues is smaller than for both academic presentations (t=-5.51,$p<0.001$) and read speech (t=-3.25, $p<0.05$). We also found that the /g/ intensity difference in simulated public speech is smaller than in read speech (t=-3.22,$p<0.05$). The other comparisons are not significant. As shown by the right panel in Figure 3, the relationships among the speech styles in /d/ and /g/ display, to a great extent, what we expected and the intensity differences across speech styles are relatively consistent. The intensity difference is smaller in more casual speech styles (i.e., dialogues and simulated public speech) as compared to careful speech (i.e., academic presentation and read speech). Our findings here support what we discussed above that the intensity difference measure is more useful for voiced stops than for nasals.

4 Conclusion

In the present study, we used a large-scale spontaneous speech corpus to examine the phonetic variability of nasals and voiced stops in Japanese. We attempted to account for how the variation and reduction of nasals and voiced stops occur across speech styles by measuring the duration and the intensity difference of target segments. We hypothesized that the shortest duration and the smallest intensity difference would be observed in dialogues and the longest duration and the largest intensity difference in read speech. Unlike what we predicted, the phonetic variability exhibited complex patterns across both phonemes and speech styles. As a result, our findings revealed a few important aspects of the phonetic variability and the effect of speech styles. First, the way in which speakers reduce segments is not consistent across both speech style and phoneme. That is, the way speakers implement reduction is variable depending not only on speech style but phoneme. Second, the relationships among the speech styles in closure and release durations vary. In other words, the way in which speech styles influence closure and release durations differ. Third, the intensity difference measure is more useful for voiced stops than for nasals. Further research is needed

to investigate the instances where the segment is deleted and the segments that are realized as different phonemes (e.g., /d/ → [ɾ]).

References

Arai, T., Warner, N., and Greenberg, S. (2007). "Analysis of spontaneous Japanese in a multi-language telephone-speech corpus". *Acoustical Science and Technology*, 28:1, 46–48.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2017). "Package lme4". *R package version*, 1.1-13.

Boersma, P. and Weenink, D. (2016). "Praat: doing phonetics by computer [Computer program]".

Brand, S. and Ernestus, M. (2015). "Reduction of obstruent-liquid-schwa clusters in casual french". In *the 18th International Congress of Phonetic Sciences (ICPhS 2015)*.

Ernestus, M., Baayen, H., and Schreuder, R. (2002). "The recognition of reduced word forms". *Brain and Language*, 81:1–3, 162 – 173.

Ernestus, M. and Warner, N. (2011). "An introduction to reduced pronunciation variants". *Journal of Phonetics*, 39:3, 253–260.

Johnson, K. (2004). "Massive reduction in conversational American English". In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.

Kohler, K. J. (1990). "Segmental reduction in connected speech in german: Phonological facts and phonetic explanations". *Speech production and speech modelling*, 55, 69–92.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). "Package lmerTest". *R package version*, 2.

Lennes, M., Alarotu, N., and Vainio, M. (2001). "Is the phonetic quality of unaccented words unpredictable? an example from spontaneous finnish". *Journal of the International Phonetic Association*, 31:1, 127–138.

Lenth, R. (2017). "Package lsmeans". *R package version*, 2.26-3.

Maekawa, K. (2003). "Corpus of Spontaneous Japanese: Its design and evaluation". In *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*.

Maekawa, K. (2005). "Toward a pronunciation dictionary of Japanese: Analysis of CSJ". In *Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2005)*, (pp. 43–48).

R Core Team (2017). "R: A language and environment for statistical computing". *R Foundation for Statistical Computing. Vienna, Austria*.

Torsten Hothorn, Frank Bretz, P. W. R. M. H. A. S. S. S. (2016). "Package multcomp". *R package version*, 1.4-6.

Tucker, B. V. (2007). *Spoken word recognition of the reduced American English Flap*. phdthesis, The University of Arizona.

Tucker, B. V. (2011). "The effect of reduction on the processing of flaps and /g/ in isolated words". *Journal of Phonetics*, 39:3, 312–318.

Warner, N. and Tucker, B. V. (2011). "Phonetic variability of stops and flaps in spontaneous and careful speech". *The Journal of the Acoustical Society of America*, 130:3, 1606–1617.

A5

壱岐勝本方言のアクセント体系とアクセント単位[1]

与那国町教育委員会　　中澤 光平

１．はじめに

本発表は，長崎県壱岐島の勝本浦で話されている勝本方言のアクセント体系およびアクセント単位について，発表者の調査データをもとに考察することを目的とする。

２．壱岐市勝本及び勝本方言について

壱岐島は, 九州北方の玄界灘にある島で, 九州と対馬の中間に位置する。現在は島全体が壱岐市の 1 市体制だが, 2004 年に合併するまでは戦後長らく芦辺町（北東），石田町（南東），郷ノ浦町（南西）および勝本町（北西）の 4 町に分かれていた。

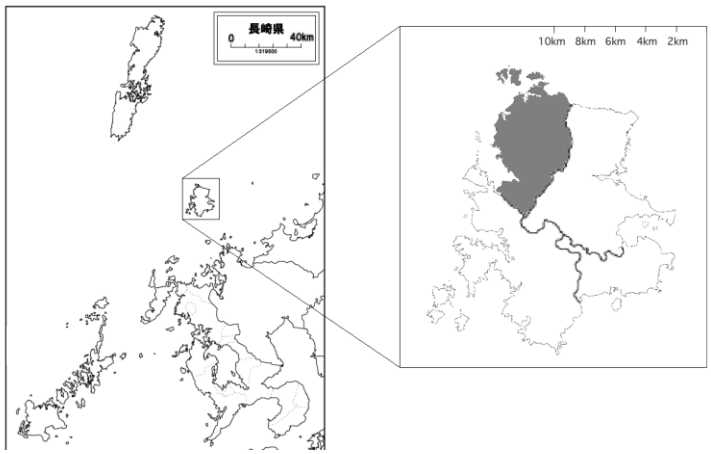


図１　壱岐島および勝本町（灰色部分）の位置

壱岐市は長崎県に属するが,「交通・交易は古くから福岡との間に栄えている」（岡野信子 1983: 145）ため，壱岐方言はむしろ筑前に近いと考えられる。地元の研究家による鳥巣修 (2008) も,「壱岐は、江戸期に平戸藩に属していたが、交通・交易も福岡・佐賀との間が盛んであったため、方言の様々な面で長崎本土と違い旧筑前や旧豊前に似たところがある」 (p.10) と述べる。また，壱岐方言内部の地域差について，岡野 (1983) は,「地域差がさほど著しくないが、南東部の石田町の方言は概して古風であり、北部の勝本浦の方言には新化が認められる」として，勝本浦（勝本町の漁村集落）の方言に特色があることを述べる。勝本浦を含む勝本町は，特にアクセントの点で他の三町（ただし郷ノ浦北部の沼津は勝本型）と大きく異なることが，先行研究で指摘されている（平山輝男 1951，金田一春彦 1954，岡野 1983）。

３．先行研究

勝本方言を含む壱岐方言のアクセントに関する先行研究には平山 (1951), 金田一 (1954), 岡野 (1983), 池田史子 (2004) などがある。壱岐方言のアクセントに関する先行研究の記述をまとめると次のようになる。

[1] 本研究は国立国語研究所「対照言語学の観点から見た日本語の音声と文法」（プロジェクト代表者：窪薗晴夫）による「若手研究者のためのアクセント調査旅費」プログラムの支援を受けている。

(1) 平山 (1951) による壱岐方言の名詞アクセント

一音節名詞	中・南部	H-L / H-L（-が等 / -の,ん）例：柄(エ)，蚊(カ)，子(コ)	
		H-L / L-L (H-H?)（同上）例：絵(エ)，木(キ)，粉(コ)	
二音節名詞	中・南部	LH / LH-L / LH-H（Ø(ゼロ) / -が等 / -の,ん）例：飴(アメ)，米(コメ)，鼻(ハナ)・花(ハナ)	
		HL / LH-L / HL-L（同上）例：雨(アメ)，傘(カサ)，喉(ノド)，猫(ネコ)	
三音節名詞	中・南部	HLL	例：烏(カラス)
		LHL	例：心(ココロ)
		LHH(-L)	例：頭(アタマ)
四音節名詞	中・南部	HLLL	例：あくる日，くちばし
		LHLL	例：あかぎれ，かまぼこ
		LHHL	例：鉢巻き，雨傘
		LHHH(-L)	例：薙刀，小刀

H: 高い拍。L: 低い拍。以下同様。

母音の無声化は省略した。平山 (1951) の北部は勝本型を，中・南部はそれ以外の型を表す。これを見ると，一音節名詞（1 拍語）には 2 つの型（一部で中和），二音節名詞（2 拍語）には 2 つの型（一部で中和），三音節名詞（3 拍語）には 3 つの型，四音節名詞（4 拍語）には 4 つの型があると認めているようである。北部と中・南部との違いは 2 拍語の類別語彙の統合の仕方にあると述べている（類別語彙の統合については後述）。

(2) 金田一 (1954) による壱岐方言の名詞アクセント

壱岐主流方言	LH ~ LH-L（-が）	例：風，牛，音，石
	LH ~ LH-H / LH-L（-の / -が）	例：皮，足
	HL ~ HL-L（-が）	例：海，箸，秋，猿
壱岐西北部方言	HL ~ HL-L（-が）	例：風，牛，音，石
	HH ~ HH-H / HH-L（-の / -が）	例：皮，足
	HL ~ HL-L（-が）	例：海，箸，秋，猿

金田一 (1954) は 2 拍語のみの考察に限るが，壱岐主流方言に 3 つの型を認め，西北部方言（勝本型に相当）に 2 つの型を認めている。

(3) 岡野 (1983) による壱岐方言の名詞アクセント

一音節名詞 H-L　二音節名詞 LH, HL　三音節名詞 LHH, LHL, HLL

岡野 (1983) は，各型の所属語彙こそ違うが，石田町も勝本浦も同じ体系と述べている。

(4) 池田 (2004) による壱岐勝本方言の名詞アクセント

1 拍名詞 H / H-L(…) / H-H(…) (L-H(…)?)

2 拍名詞 HL / HL-L(…) / HL-L(…)
LH / LH-L(…) / LH-H(…)

3 拍名詞 LHL / LHL-L(…) / LHL-L(…)
HLL / HLL-L(…) / HLL-L(…)
LHH / LHH-L(…) / LHH-H(…)

4 拍名詞 HLLL / HLLL-L(…) / HLLL-L(…)
LHLL / LHLL-L(…) / LHLL-L(…)
LHHH / LHHH-L(…) / LHHH-H(…)

5 拍名詞 LHHLL/LHHLL-L(…)/LHHLL-L(…)
LHHHH/LHHHH-L(…)/LHHHH-H(…)

(Ø（ゼロ） / -が,で,に,から,より等 / -ん,だけ,ばかり等)

池田 (2004) は本発表と同じく勝本方言に限定した上で，5 拍語までの名詞アクセントについて記述している。池田 (2004) によれば，名詞の大部分はアクセントの下がり目が後ろから 1 拍目にあるグループと（－①と表記），アクセントの下がり目が後ろから 3 拍目にあるグループと（－③と表記）に分けることができる（1 拍語は 1 つの型のみ）。また，「が」のように常に低く接続する助詞（「助詞 A グループ」）と「ん」（「の」の転）のように直前の拍と同じ高さに接続する助詞（「助詞 B グループ」）がある。

いずれの先行研究も類別語彙[2]と類の統合について言及している。

表 1　先行研究による壱岐勝本方言の類の統合

	1 拍名詞	2 拍名詞	3 拍名詞
平山 (1951)	1·2/3?（中・南部）	1·2·4·5/3	
金田一 (1954)		1·2·4·5/3	
岡野 (1983)	1·2·3?	1·2·4·5/3	1·6·7/3/4?
池田 (2004)	1·2·3 (< *1·2/3)	1·2·4·5/3	1·2·3·5·6·7/4

x·y: x と y が統合している。x/y: x と y が区別される。斜線: 先行研究で言及無し。

先行研究で一致しているのは，2 拍名詞の類の統合を「1·2·4·5/3」と考えることである。池田 (2004) によれば，勝本方言はあたかも「二型アクセント」の様相を呈する。3 拍語の LHL は「単衣」の 1 語のみで，しかも HLL とで「揺れているように感じられる」という。4 拍語には HLLL が 8 語あるが，LHLL の 56 語，LHHH の 39 語と比べると少数と言える。

一方，岡野 (1983) は勝本方言にも LHL を 17 語認めている。HLL の 43 語および LHH の 36 語と比べれば少数であるものの，決して無視できない数である。平山 (1951) は 3 拍語について，中・南部との違いには言及していない。

[2] 過去の文献，ならびに現代諸方言の考察から，古い日本語において同じアクセントを持っていたと推定される語彙をグループにまとめたもの。金田一 (1974) では，名詞について以下の類を立てる（p.62–73）。1 拍名詞 : 1 類～3 類，2 拍名詞 : 1 類～第 5 類，3 拍名詞 : 「形」類（1 類），「小豆」類（2 類），「頭」類（4 類），「命」類（5 類），「兎」類（6 類），「兜」類（7 類）。金田一 (1974) では立てられていないが，「力，二十歳，…」などを「二十歳」類（3 類）として認める場合もある。

４．勝本方言のアクセント体系

LHL の違いのように，先行研究で一致しない部分は，あるいは歴史的な変化を示しているのかもしれない。このような点を明らかにするため，発表者は壱岐での現地調査を行った。本発表では発表者の調査データをもとに勝本方言のアクセント体系について記述する。

４．１　調査概要

調査は 2013 年 9 月と 2017 年 1 月に実施した。2013 年の調査は単独形での調査が中心で，2017 年の調査は主に助詞付きの形での調査を行った。話者は 2013 年と 2017 年で異なる。調査方法は発表者が作成した調査票を読み上げていただく形で行った。各話者とも調査時間は 1 時間～2 時間程度である。

４．２　名詞のアクセント体系

下に，発表者による壱岐勝本方言の名詞アクセントの調査結果を示した。

表２　壱岐勝本方言の名詞のアクセント体系

型＼拍数	1 拍語	2 拍語	3 拍語	4 拍語
1 型	H, H-L	HL, HL-L	HLL, HLL-L	(HLLL, HLLL-L)
2 型		LH, LH-L / LH-H	(LHL, LHL-L)	LHLL, LHLL-L
3 型			LHH, LHH-L / LHH-H	(LHHL, LHHL-L)
4 型				LHHH, LHHH-L

（‘-’の後は 1 拍助詞「-が / -の」の音調）

n 型は n 拍（n モーラ）目の後に下がり目がある（＝n 拍目に下げ核がある）ことを示す。ただし－①型では，池田 (2004) の助詞 B グループが後続する場合下がり目が消える。

()に示した型は，少数の上に個人差があり，また個人の中でも安定しているものは少ない。

(5)　a. LHL で出た語　　欠伸(ｱｸﾋﾞ)，砂糖(ｻﾄｰ)，二重(ﾌﾀｴ)，双子(ﾌﾀｺﾞ)，二つ(ﾌﾀﾂ)

b. HLLL で出た語　　居眠り(ｲﾈﾑﾘ)，猪(ｲﾉｼｼ)，金槌(ｶﾅｽﾞﾁ)，楠(ｸｽﾉｷ)，嘴(ｸﾁﾊﾞｼ)，座布団(ｻﾞﾌﾞﾄﾝ)，魂(ﾀﾏｼｰ)，手袋(ﾃﾌﾞｸﾛ)，友達(ﾄﾓﾀﾞﾁ)，鶏(ﾆﾜﾄﾘ)，三日月(ﾐｶｽﾞｷ)

c. LHHL で出た語　　此間(ｺﾅｲﾀﾞ)，筍(ﾀｹﾉｺ)，物陰(ﾓﾉｶｹﾞ)

HLLL はある程度の語数が見られるが，LHL と LHHL はほとんど確認できなかった。

このように，壱岐勝本方言のアクセントは，類別体系や共時的アクセント型の偏りから，名詞に関して概ね 2 つの型の区別（－①と－③）しかないように見えるが，調査語数を増やすと－①と－③に収まらない語が出てくることから，核の位置が弁別的な体系と位置付

けられる。ただし無核型を欠くため，1 拍語に 1 種類，2 拍語に 2 種類，3 拍語に 3 種類，…と，n 拍語に n 種類のアクセント型があることになる。

４．３　不定語を含む文のアクセント

壱岐勝本方言のアクセント体系は，付属語を含む文節がそのアクセント単位となり，名詞の核の位置と助詞の種類でアクセント型が決まることが分かった。一方で，一部の環境ではこの原則から外れることが，今回の調査で明らかになった。例外的なアクセント型が観察される環境の一つとして，「何，誰」などの疑問詞が不定語となる場合，不定語を含む文節全体が，後続する助詞が助詞 B グループでなくとも下がり目を失う。

(6) ナニ LH（何），ナニガ アル LHL HL（何がある），ナンバ タベチモ ヨカヨ LHH HHHL HLL(↑)（何を食べても良いよ）※(↑)は上昇調イントネーションを表す。以下同じ。

「バ」は「ガ」と同様に助詞 A グループなので通常は低く付く（例：イヌバ LHL（犬を））が，(6)ではナンバ LHH という予想（LHL）と異なる音調を取る。

疑問詞が助詞「モ」を伴い不定詞と解釈されるとき，「モ」が疑問詞を含む文節に含まれる場合は，「モ」の直前に下がり目が来る。「モ」が疑問詞を含む文節に含まれない場合は，疑問詞を含む文節が平板化する。

(7) イツ LH（何時），イツマデモ アル LHHHL HL（何時までもある），イツマデ タッチモ オワラン LHHH HHHL HHLL（何時まで経っても終わらない）cf. ナミマデ LHLL（波まで）

また，「モ」が疑問詞とは別の文節にある場合，「モ」の直前まで平板化することがある。

(8) ナンバ コーチモ ヨカヨ LHH HHHL HLL(↑)（何を買っても良いよ）cf. コーチ HLL

５．考察

名詞のアクセント調査から，勝本方言は非常に偏りがあるものの，核の位置が弁別的な体系と解釈されることを示した。ただし無核型を欠き，助詞 B グループが後続する場合に限り，－①型は下がり目が消える。しかし，不定語を含む文では，助詞 A グループでもナンバ LHH（何を）のように下がり目が消え，無核型を認める必要があるように思われる。

この現象については，例外的に無核型を認めるのではなく，この場合も「モ」の直前に下がり目が生じることから，疑問詞から「モ」の直前まで（[疑問詞…]_モ）が，あたかも一語として扱われるものと解釈したい。(9)では[ナンバ コーチ]までを－①型の語と見なす。

(9) [ナンバ コーチ]モ ヨカヨ LHH HHHL HLL(↑)（何を買っても良いよ）

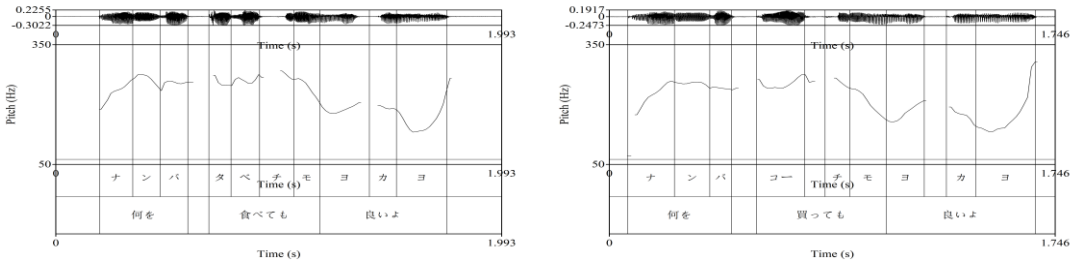


図２　「ナンバタベチモヨカヨ/ナンバコーチモヨカヨ」のアクセント

このような単位を認めることはアドホックにも思われるが，佐藤久美子 (2016) によれば，長崎市方言でも同様の環境において，不定語（疑問詞）から「モ」の直前までが「音韻的な語となる」と分析している。ただし，佐藤 (2016) ではこの「音韻的な語」と「モ」が複合語となるとしているが，壱岐勝本方言では「モ」は助詞 A グループの一般的な振舞いを示すものと解釈される。ただし, (10)のような例は一語化の解釈ができず, 課題として残る。

(10) ダレガ イチバデ リンゴバ コーチモ ヨカヨ LHH LHHL HLLL HHHL HLH(↑)（誰が市場でリンゴを買っても良いよ）（イチバデ，リンゴバが元のまま，ダレガだけ平板化）

６．まとめと課題

壱岐勝本方言のアクセントは，長崎市方言にも想定されるような「音韻的な語」を認めることで，無核型の欠けた核の位置が弁別的な体系と解釈できるが，それだけでは説明できない例も残った。また，本発表では扱わなかったが，旧勝本町でも湯本地区など勝本浦以外では LHL などがある程度見られた。近隣方言の調査により，名詞アクセントの著しい偏りが歴史的に生じたことを明らかにできる可能性がある。

謝辞

話者の皆様および話者をご紹介くださった鳥巣修様にお礼申し上げます。

参考文献

池田史子 (2004) 「壱岐勝本方言の複合名詞アクセント : 後部要素が 1 拍名詞の場合の複合規則仮説」『山口県立大学國際文化學部紀要』10: 21–27.
岡野信子 (1983) 「壱岐・対馬の方言」『講座方言学９　九州地方の方言』: 143–171.
金田一春彦 (1954) 「対馬附壱岐のアクセントの地位―九州諸方言のアクセントの対立はどうしてできたか」九学会連合対馬共同調査委員会『対馬の自然と文化』: 347–373.
――― (1974) 『国語アクセントの史的研究』pp.318, 東京, 塙書房.
――― (2005) 『金田一春彦著作集　第七巻』pp.754, 町田, 玉川大学出版部.
佐藤久美子 (2016) 「長崎市方言における複合法則の適用について」「対照言語学の観点から見た日本語の音声と文法」研究発表会, 発表資料.
鳥巣修 (2008) 『きーちみ　しゃべっちみ―壱岐島の方言事例集―』: 長崎.
平山輝男 (1951) 「壹岐對馬兩方言の音調に就いて」『音声の研究』7: 213–228.

A6

東海地域西部方言のアクセントにおける「おそあがり」の音声実現の特徴

吉田 健二（愛知淑徳大学文学部）
kenjiyo.work@gmail.com

1. 問題点

本稿の目的は，名古屋方言アクセントの音声実現上の特徴といわれる，語頭のピッチ上昇のおくれ（以下，「おそあがり」）の音声実現上の特徴を，関東の東京式アクセントとの対比によりあきらかにすることである．東海地域と関東地域，各 5 名の話者の発話データの定量的分析にもとづき，ピッチ動態のどこに関東とことなる特徴が顕著にあらわれるかを検討する．

名古屋を中心とした東海地域には，内輪式アクセント（上野 1987：44）が分布し，関東などの中輪式とアクセント体系・音声実現の両面で共通する点がおおい．いっぽう，東海地方に特有の音声実現上の特徴として，(1)にしめすような，おそあがりの傾向が指摘されている（水谷修 1960：9 より．ピッチの値は拍単位の L＝低 / H＝高でしめす）．

(1)　（東京）（名古屋）　（東京）（名古屋）　（東京）　（名古屋）
桜　LHH　LLH　玉ねぎ　LHHL　LLHL　東山　LHHHH　LLHHH

3 拍以上の語で語頭の上昇が東京などにくらべて 1 拍分おくれ，2ー3 拍めで実現するということである．水谷（1960）は自発発話にみられるおそあがりの例もしめしており，筆者も東海圏でおそあがりと聴かれる発話をしばしば耳にする．この傾向は，関東方言とことなる，東海方言固有の特徴といえるだろうか．また，ちがいがあるとすれば，具体的にはピッチ実現のどのような特徴にそれがあらわれるのだろうか．この問題を検討するための端緒としておこなった音声データの分析結果と，えられた知見を報告する．

2. 方法

2.1. 話者

東海方言の話者は，愛知淑徳大学文学部「国語学演習」で実施した言語調査でお会いした 5 名で（吉田健二・他 2016, 2017），表 1 のとおり，愛知県最西端の愛西市旧八開村，立田村，弥富市と，木曽川をへだてて隣接する岐阜県海津市のかたがたである．調査は 2015 年 9 月と 2016 年 9 月に実施しており，内輪式アクセントとみられる話者は 30 名いた．この 5 名はそのなかから，聴覚印象からおそあがりの傾向をもつとおもわれる話者を中心に，代表としてえらんだ．他の話者についてもおなじ分析を実行中である．関東方言の話者は，東海の話者とおよそ同年代の話者を中心に，5 名におねがいした．表 1 の生育地は 6-15 歳のときにいた場所である．一時的に他の地域に居住した方が 4 名いるが，この期間については，全員，表 1 にしめす地域のみですごしている．

表 1: 話者情報

略称	生年	性別	生育地	地域
yatomi30	1987	男	愛知県弥富市	東海
hachikai30	1986	女	愛知県愛西市（旧八開村）	東海
tatsuta30	1987	女	愛知県愛西市（旧立田村）	東海
tatsuta40	1971	男	愛知県愛西市（旧立田村）	東海
kaizu40	1971	男	岐阜県海津市（旧平田町）	東海
tokyo20	1996	女	東京都練馬区	関東
tachikawa30	1978	女	東京都立川市	関東
saitama50	1965	女	埼玉県狭山市	関東
kawasaki40	1977	女	神奈川県川崎市	関東
chiba30	1982	女	千葉県千葉市	関東

2.2. 音声データ

分析対象の音声データは，(2)のような「名詞＋の＋名詞＋断定辞」という構造の短文の読み上げ発話によるものである．文末の断定辞は，なじみのあるものにおきかえてもらった（東海は「やん・じゃんか・だがや」，関東は「じゃん」）．三重の中央式アクセントにおけるダウンステップパターン検討のためのコーパスで（吉田・他 2016），先行語（ひとつめの名詞）は，中井（2002）で京都の 16 人全員が高起無核 / 低起無核 / 低起 2 核で発音した 3 拍名詞，後続語（ふたつめの名詞，以下「実験語」）はおなじく全員が高起無核 / 低起無核で発音した 3ー5 拍名詞である．後続語 16 語を京都の音調型ごとに(3)にしめす．

(2) 実験文の例

田舎の祭りやん　　田舎の野良猫やん　　お金の悩みごとやん

(3) 実験語（3 拍 : 2 語，4 拍 : 5 語，5 拍 : 1 語）

高起無核 祭り，名前，乗り物，ものまね，日本間，二枚目，難問，悩み事

低起無核 煮豆，マンガ，生ハム，人形，野良猫，持ち逃げ，人間，旦那さん

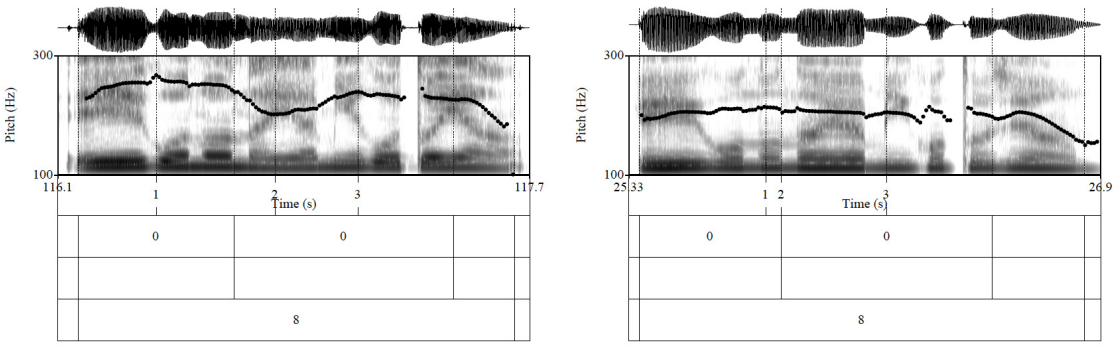
実験語 16 語を先行語の 3 種類の音調型条件においたので，実験文はぜんぶで 48 種類．発話は各文 1 回．10 名とも実験語のほとんどを無核型（平板型）で発音したので，ひとりにつき平板型のデータが 40ー46 例えられた．

2.3. 音声データのアノテーションとデータ処理

ピッチ動態の定量分析のための代表値を抽出するため，各音声データに，(4)にしめすアノテーションをほどこした（以下，音声の基本周波数を fo と略称）．図 1 に 2 例をしめす．

(4) 先行語内の fo ピーク : P1（図 1 の 1）
実験語初頭の fo の谷 : V1（図 1 の 2）
実験語内の fo ピーク : P2（図 1 の 3）

ただし，関東の話者および東海の一部（次節参照）に，実験語の初頭にピッチの谷がなく，先行語から連続してピッチが単調に下降するケースがみられた（図 1 右）．こういうケースでは，V1, P2 を確定する基準がえられない．明瞭な谷があるケースとの比較にあたり，とくに P2 のタイミングについて，かたよりが最小になる方法として，かりに先行語と後続語の境界時点を V1，後続語の持続時間の中間点を P2 とした．



hachikai30（東海）　　kawasaki40（関東）

図 1: 音声データのアノテーションの例「英語の悩みごとやん」

この方法により確定した P1, V1, P2 の 3 時点について，実現時間と fo 値を抽出した．fo レンジがおおきくことなる話者間の比較も可能にするため，fo 値は各話者の平均値を基準とした標準得点（z-score）に変換した．

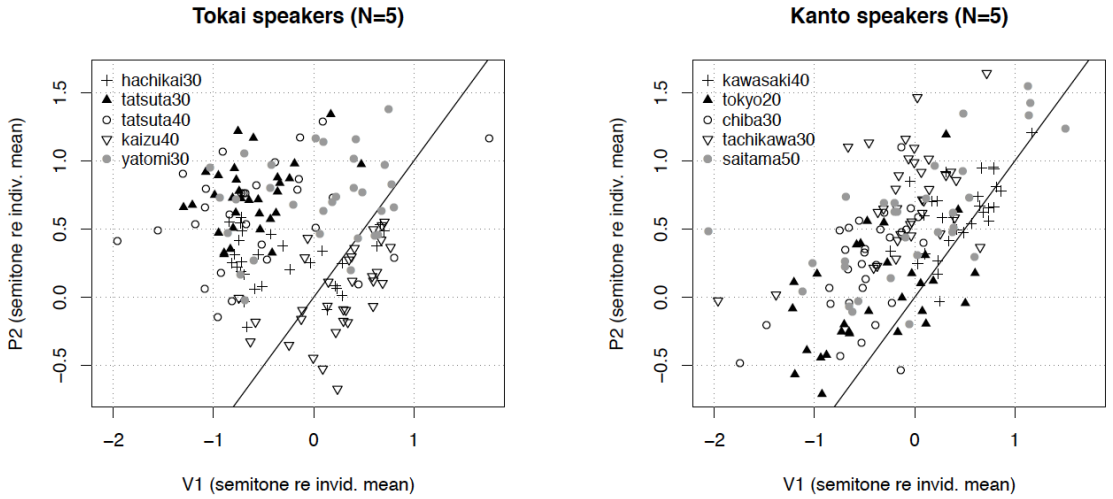
3. 結果

3.1. V1 から P2 にかけての fo 値のうごき

以下では，検討の対象を先行語・実験語ともに平板型のケースにしぼる．先行語も平板型で発音されたケースがおおかったので，各話者最低 29 例（29ー35 例）えられている．図 2 に V1×P2 の全ケースの fo 値を話者グループ別にしめす．右の関東の 5 名は，個人によるバラツキが比較的ちいさい．総じて，V1＝P2 の補助線にそって，その左上にデータがならぶ．先行語と実験語のあいだのピッチの谷（V1）から，実験語のピーク（P2）にかけてちいさな上昇があるケースが大半をしめる，ということである．V1＝P2 の補助線付近および右下に位置するケースは，2.3 節でのべた，先行語から実験語にかけてピッチが単調下降をしめすケースの存在を反映する．

これにたいして，左の東海 5 名は，話者間・話者内のバラツキがおおきい．2.3 節でのべたとおり，yatomi30, kaizu40 に前述の先行語から実験語にかけてピッチが単調下降するケースがおおく，V1＝P2 の補助線付近にデータが集中する．このタイプを関東で典型的にみられるピッチ動態パターンとかんがえ，かりに「**関東タイプ**」とよぶ．関東タイプは，上記 2 名だけでなく，数はすくないものの hachikai30, tatsuta40 にもみられる．いっぽう，東

海の話者には V1＝P2 の補助線から左上にややはなれた位置のデータもおおい．先行語と実験語のあいだのピッチの谷（V1）がふかく，そこから実験語のピーク（P2）にかけてのピッチ上昇がおおきいことをしめす．このタイプを東海で典型的にみられるピッチ動態として，「**東海タイプ**」とよぶ．このタイプがおおいのは tatsuta30 と tatsuta40 だが，yatomi30, hachikai30 にも数はへるものの，このタイプがみられる．東海の話者には，関東タイプと東海タイプの音声実現タイプが混在している，ということである．

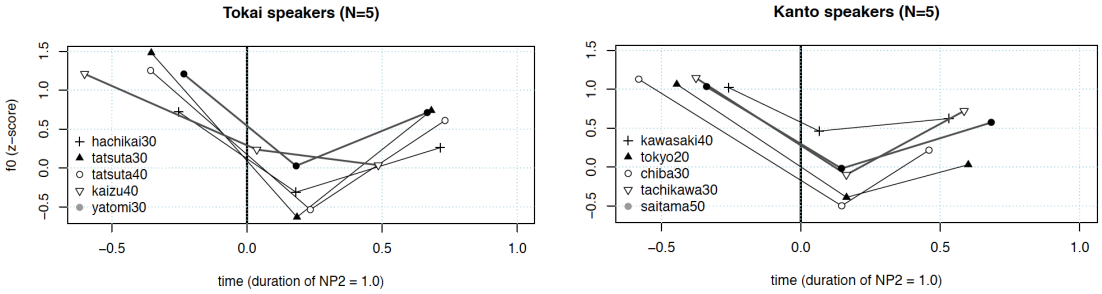


東海　　関東

図 2: V1（横軸）x P2（縦軸）の fo 値（話者内 z-score）．斜線は V1＝V2.

3.2.　P1・V1・P2 の実現タイミング

つぎに，各 fo 代表値が実現したタイミングもあわせてピッチ動態全体の傾向を検討する．話速のちがいを捨象して比較できるよう，実験語の持続時間を 1.0，開始時間を 0.0 として，P1，V1，P2 の実現時間を相対的時間に変換した．この値からもとめた，3 つの代表値のタイミングと fo 値の話者ごとの平均値を図 3 にしめす．東海の話者は，前節で**関東タイプ**がおおいことを確認した kaizu40 をのぞき，P1 から V1 にかけての下降がおおきく，ふたたび P2 にかけておおきく上昇する．結果として，先行語から実験語にかけてのピッチの谷がふかくなる傾向がある．これが**東海タイプ**のピッチ実現上の特徴とかんがえられそうである．これにたいして，関東の話者は，下降も上昇もちいさめで，ピッチの谷があさい．いっぽう，P1, V1, P2 の実現タイミングには，東海・関東におおきな差はみられないが，やはり kaizu40 をのぞく東海の話者のばあい，P2 のタイミングがややおそい傾向がある．この点については，次節でさらに検討する．



東海　　関東

図 3: P1, V1, P2 のタイミングと fo 値の平均値（実験語の持続時間＝1.0 / 開始時間＝0.0）

3.3. fo 値とタイミングによる話者のことなりのパターン

前節までででみた関東と東海のちがいは，1 節で提起した「東海方言固有」といえるだろうか．このことを検討するため，3 つの fo 代表値とそのタイミング，計 6 つの測定値を変数とした主成分分析を実行した．第 1 成分（寄与率 31%）と第 2 成分（寄与率 27%）の biplot を図 4 にしめす．太字が 10 名の話者，イタリックが 6 つの測定値のふたつの成分の重みによる位置である．まず話者の位置をみる．kaizu40 をのぞく東海の話者は，第 1 成分が正の領域に分布している．いっぽう関東の話者は第 1 成分で 0 付近または負の領域に分布している．両者が明確に分離していないのは，3.1 節でみたように，東海の話者に東海タイプと関東タイプの音声実現が混在することによるとかんがえられる．その傾向つよかった yatomi30 は，第 1 成分で 0 付近に位置しており，関東の話者とちかいことがとらえられている．

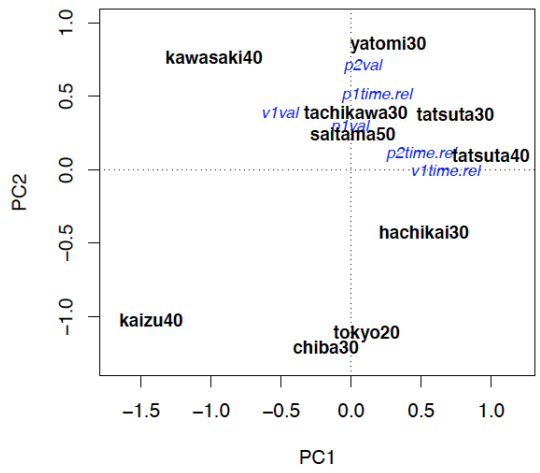


図 4: 6 つの測定値による主成分分析の結果（第 1 成分 × 第 2 成分）

つぎに，測定値の位置をみる．上の検討から関東タイプ・東海タイプのちがいをとらえたのは第 1 成分だとかんがえられるが，この重みが正または負におおきい音声特徴は，おおきさの順に(a)P2 の実現時間（東海がおそい），(b)V1 の実現時間（東海がおそい），(c)V1 の fo 値（東海がひくい）である．これらは，3.1, 3.2 節で検討した特徴であり，「東海タイプ」固有のピッチ動態上の特徴とかんがえられそうである．

4. まとめと課題

東海アクセントの「おそあがり」傾向の実態を検討した．結果は東海固有の音声実現上の特徴をうかびあがらせたが，同時に，あきらかに関東側に分類される話者（kaizu40）や，東海固有の音声特徴とともに，共通語的な音声実現もあわせもつ個人が存在することもしめした．東海地域にも当然，共通語化の波が押し寄せており，この結果はそれを反映しているとおもわれるが，水谷（1960）の報告から半世紀を経てなお，今回対象とした比較的わかい層に，地域固有の音声実現の特徴もみいだされることが確認できた．

「おそあがり」の音声実現上の特徴のひとつに，連続する平板型のあいだのピッチの谷のふかさがあった．東京方言の語頭のピッチ上昇について，弱化したり実現しなかったりすることもあるが，アクセント単位の特徴としてつねに存在するというみかたがあるが（郡2004），東海方言アクセントの音声実現にはその傾向がよりつよいとみることができるだろうか．今後の課題としたい．また，これと関連して興味ぶかくおもわれるのが，この地域に，一部の複合名詞の前部・後部要素のあいだにピッチの下降・上昇をおく傾向がみられることである（「英作文」「新横浜」の HHLHLL 型，「反復横跳び」の HLLLLHHH 型など：吉田・他 2017：232-237 で報告）．今回みた，句レベルの連接における谷のふかさが，語レベルの連接（複合）における，ふたつの要素の韻律上の独立性のつよさと関連する可能性，たとえば，両者ともに東海方言に一般的な韻律上の特徴の反映である，というような可能性はないだろうか．これも今後の検討課題としたい．

謝辞　音声データを提供してくださったみなさまに感謝もうしあげます．また，名古屋音声研究会の研究例会上で，重要なコメントをいただきました．この研究は，愛知淑徳大学の「学外教育等活動」予算および，文部科学省科学研究費助成金（「日本語諸方言のプロソディーとプロソディー体系の類型」研究代表者：窪薗晴夫，課題番号 26244022）による助成を受けています．

参考文献

上野善道 (1987) 「日本本土諸方言アクセントの系譜と分布(2)」『日本学士院紀要』42:1, 15-70.

郡史郎 (2004) 「東京アクセントの特徴再考－語頭の上昇の扱いについて－」『国語学』55:2, 16-31.

中井幸比古 (2002) 『京阪系アクセント辞典』東京：勉誠出版

水谷修（1960）「名古屋アクセントの一特質（前半）」『音声学会会報』102, 8-10.

吉田健二・他 (2016) 「三重・愛知県境方言における方言の接触と変容」『愛知淑徳大学国語国文』39, 218-250.

吉田健二・他 (2017) 「愛知・岐阜・三重県境地域における言語使用と言語意識」『愛知淑徳大学国語国文』40, 207-242.

B1

音象徴の抽象性：赤ちゃん用オムツのネーミングにおける唇音

熊谷 学而（国立国語研究所）川原 繁人（慶應義塾大学言語文化研究所）
gakuji-kumagai@ninjal.ac.jp, kawahara@icl.keio.ac.jp

1. はじめに

音象徴(sound symbolism)とは、「ある特定の音が、あるイメージ（意味）に結びつく」という現象で、音声学・心理学・認知科学で活発に研究されている分野である (e.g., Blasi et al. 2016; Dingemanse et al. 2015; Hinton et al. 2006). 近年の音象徴研究の草分け的存在である Sapir (1929)は、英語話者を対象に「大きいテーブルと小さいテーブルを示す単語が 2 つあった場合、どちらが“mal”で、どちらが“mil”か」と尋ね、多くの話者が「大きいテーブル＝“mal”」「小さいテーブル＝“mil”」と選ぶという結果を得た．この結果は、英語話者にとって「[a]＝大きい、[i]＝小さい」というつながりが存在することを示唆している．発表者らは、大学の言語学や音声学の入門授業で、同じ実験を履修生相手に試すことも多いが、日本人話者を相手にしても同じような結果が得られることが多い．また、このような音象徴の実験を授業中に行うと、学生の興味も惹きやすい．近年、音象徴は「音声学で使用される概念の導入」としても有効であるという見方もあり（川原 2015, 2017a, b; Kawahara et al. 2016; Kawahara & Kumagai 2017）、以下で示す本研究の成果もそのような試みにも貢献する．

本研究は、赤ちゃん用オムツの名前について、音象徴の観点から実験を行い、考察する．日本で市販されている赤ちゃん用オムツには、「パンパース (**p**an**p**aasu)」「メリーズ (**m**eriizu)」「ムーニーマン (**m**uuni**i****m**an)」「マミーポコ (**m**a**m**i**i****p**oko)」など、名前に両唇音である[p, m]が含まれているものが多い．両唇音は、喃語の中に多く観察され、「パパ (papa)」や「ママ (mama)」のような単語に使われることからも分かるとおり、赤ちゃんが最初に獲得する子音でもある (Jacobson 1941/1968). よって、音象徴の観点から考えると、「唇音＝赤ちゃん」というつながりが成り立っている可能性がある．ただし、オムツの名前に関しては、実例は上にあげたような 4 つほどであり、この音象徴的つながりが一般的な法則として成り立っているかは精査する必要がある．よって、本研究では「赤ちゃん用オムツの名付けに両唇音が多用されるか」を実験により検証する．（本稿では、以下、両唇音を簡略的に「唇音」と呼ぶ.）

2. 実験 1

2.1. 方法と手順

「唇音＝赤ちゃん用オムツの名前」というつながりが存在するかどうか検証するために、実験 1 では、5 つの唇音[p, b, m, ɸ, w]をそれぞれ語頭に含む無意味語と、対照群の無意味語を用意し、日本語母語話者に、どちらが「赤ちゃん用オムツとして相応しい名前か」を選択してもらった．表 1 に、実験に用いた 15 ペアの無意味語を示す．唇音を含む刺激では、

語頭の唇音に加えて、語中にも唇音が含まれている．一方、唇音を含まない対照群の刺激では、刺激の語頭や語中の唇音の代わりに、舌頂音 (coronal) ([t, d, n, s, j])や舌背音 (dorsal) ([k])を用いた．

実験は、Surveymonkey を利用して、オンラインで行った．被験者には、15 ペアそれぞれにおいて、どちらが赤ちゃん用オムツとして相応しいネーミングか選択してもらった．ペアの提示順序や、各ペアの刺激の提示順序は、被験者ごとにランダム化した．

表 1: 実験1で用いた刺激ペア

語頭	唇音を含む刺激			語頭	唇音を全く含まない刺激	
[p]	パラピル	**p**ara**p**iru	vs.	[t]	タラキル	tarakiru
	ペラポン	**p**era**p**on	vs.		テラコン	terakon
	ポルミン	**p**oru**m**in	vs.		トルニン	torunin
[b]	バンベル	**b**an**b**eru	vs.	[d]	ダンデル	danderu
	ベレマン	**b**ere**m**an	vs.		デレナン	derenan
	ボリッポ	**b**ori**pp**o	vs.		ドリット	doritto
[m]	マラリモ	**m**arari**m**o	vs.	[n]	ナラリノ	nararino
	メレボン	**m**ere**b**on	vs.		ネレドン	neredon
	モンパル	**m**on**p**aru	vs.		ノンタル	nontaru
[ɸ]	フレマー	**ɸ**ure**m**aa	vs.	[s]	スレラー	sureraa
	フンペル	**ɸ**un**p**eru	vs.		スンテル	sunteru
	フマーロ	**ɸ**u**m**aaro	vs.		スナーロ	sunaaro
[w]	ワポック	**w**a**p**okku	vs.	[j]	ヤトック	jatokku
	ワロモン	**w**aro**m**on	vs.		ヤロノン	jaronon
	ワボーラ	**w**a**b**oora	vs.		ヤドーラ	jadoora

2.2. 被験者及び分析

実験 1 には、日本の大学生 154 名が参加した．このうち、「日本語が母語ではない」と回答した学生 4 名や、「音象徴を研究したことがある」と回答した学生 2 名を除き、計 148 名を最終的な分析対象とした. 統計分析には、線形混交ロジスティック回帰分析 (a generalized mixed-effects logistic regression: Baayen 2008)を用い、被験者と刺激をランダム効果とした．

2.3. 結果

図 1 に、唇音を含む刺激が選ばれた割合を 5 つの唇音別に示す（但し、ɸ は f と記す）．結果、いずれの唇音に関しても「唇音を含む無意味語の方が赤ちゃん用オムツとして相応しい」と有意に選択された ($z = 7.873, p < .001$)．表 2 は、それぞれの刺激が選択された割合を示している．いずれのペアにおいても、唇音を含む刺激を選択する割合が 50%を有意に上回っている（二項分布テストを用いた事後検定で全て $p < .001$）．参加者間のばらつきを分析するために、図 2 に、刺激の 15 ペアの中で、その名前として唇音を含む刺激を選択したペア数（縦軸）とその被験者の数（横軸）を示す．「オムツの名前に相応しい名前」に唇

音を含む刺激を半数以下しか選ばなかった例外的な被験者も 18 人いたものの、大多数である 130 人が唇音を含む刺激を半数以上選んだ.

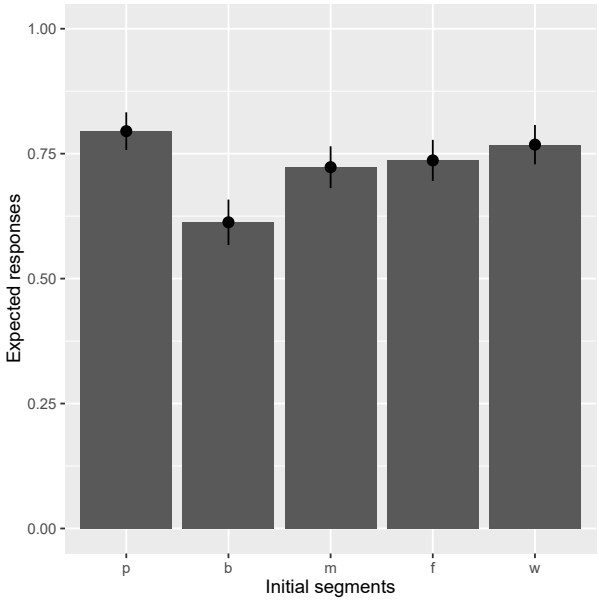


図 1: 唇音を含む刺激が選ばれた割合
（エラーバーは 95%信頼区間）

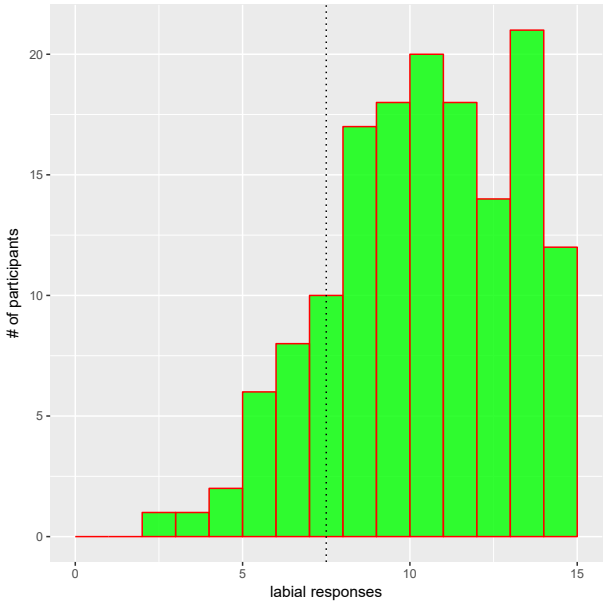


図 2: 「唇音反応数」のヒストグラム
（y 軸は被験者数）

表 2: 各刺激の選択された割合

語頭	唇音を含む刺激		割合 (%)	語頭	唇音を含まない刺激		割合 (%)
[p]	パラピル	**p**ara**p**iru	86.36	[t]	タラキル	tarakiru	13.64
	ペラポン	**p**era**p**on	72.08		テラコン	terakon	27.92
	ポルミン	**p**oru**m**in	78.57		トルニン	torunin	21.43
[b]	バンベル	**b**an**b**eru	64.94	[d]	ダンデル	danderu	35.06
	ベレマン	**b**ere**m**an	61.69		デレナン	derenan	38.31
	ボリッポ	**b**ori**pp**o	56.49		ドリット	doritto	43.51
[m]	マラリモ	**m**arari**m**o	64.29	[n]	ナラリノ	nararino	35.71
	メレボン	**m**ere**b**on	83.12		ネレドン	neredon	16.88
	モンパル	**m**on**p**aru	67.53		ノンタル	nontaru	32.47
[ɸ]	フレマー	**ɸ**ure**m**aa	91.56	[s]	スレラー	sureraa	8.44
	フンペル	**ɸ**un**p**eru	53.25		スンテル	sunteru	46.75
	フマーロ	**ɸ**u**m**aaro	75.97		スナーロ	sunaaro	24.03
[w]	ワポック	**w**a**p**okku	85.06	[j]	ヤトック	jatokku	14.94
	ワロモン	**w**aro**m**on	59.74		ヤロノン	jaronon	40.26
	ワボーラ	**w**a**b**oora	84.42		ヤドーラ	jadoora	15.58

3. 実験 2

3.1. 方法と被験者

実験 2 では、「唇音＝赤ちゃん用オムツの名前」という音象徴的つながりが生産性を持つかどうかをさらに検証するために、実験 1 に参加していない日本の女子大生 82 名に、赤ちゃん用オムツの名前を考えてもらった．比較のために、赤ちゃんのイメージが伴わない化粧品の名前も考えてもらった．それぞれ最大 3 つまで挙げてもらい、考えた名前はすべてカタカナで書くよう指示した．また、実在するオムツ名は含めないようにも指示した．実在するオムツ名やアルファベット表記をした解答は分析から除外した．

3.2. 結果

表 3 に、赤ちゃん用オムツと化粧品の名前のそれぞれに含まれている 5 つの唇音[p, b, m, ɸ, w]の割合、合計唇音数、合計子音数を示す．オムツのネーミングには、合計子音数に対して、唇音の数の割合が 48.7%であった．一方で、化粧品のネーミングには、唇音の数の割合が 19.3%しかなかった．カイ二乗検定の結果、赤ちゃん用オムツの名前には、化粧品の名前と比較して、5 つの唇音[p, b, m, ɸ, w]が有意に多く含まれていることが分かった ($\chi^2(4) = 10.72$, $p < .05$)．また、すべての唇音において、「オムツの名前に現れる割合が化粧品の名前に現れる割合よりも高い」ことも観察される．これらの結果は、実験 1 の結果と同様、「唇音＝赤ちゃん用オムツの名前」というつながりには生産性があることを示唆している．表 4 に、実験で得られた赤ちゃん用オムツのネーミングの例を示す（[w]から始まる例はなかった）．

表 3: 赤ちゃん用オムツと化粧品のそれぞれの名前に含まれる唇音の数

	オムツ	割合 (%)	化粧品	割合 (%)
[p]	106	20.4	32	5.5
[b]	31	6.0	15	2.6
[m]	63	12.1	46	7.8
[ɸ]	35	6.7	13	2.2
[w]	18	3.5	8	1.4
合計唇音数	253	48.7	114	19.3
合計子音数	519	100	587	100

表 4: 大学生が考えた、赤ちゃん用オムツの名前の例

	赤ちゃん用オムツの例
[p…]	パーニー (**p**aanii) パピー (**p**a**p**ii) ポムポム (**p**o**m**u**p**o**m**u)
[b…]	ブルン (**b**urun) ベベパン(**b**e**b**e**p**an) ベイラヴ (**b**eira**b**u)
[m…]	マパル (**m**a**p**aru) メルメル (**m**eru**m**eru) モコモン (**m**oko**m**on)
[ɸ…]	フワリー (**ɸ**uwarii) フワップ (**ɸ**uwa**pp**u) ファミー (**ɸ**a**m**ii)

4. 考察

本実験が示す興味深い点は、実在するオムツの名前に含まれる唇音[p, m]だけでなく、[b, ɸ, w]を含む名前も赤ちゃん用オムツとして相応しいことを示している点である．これは、日本語母語話者が、[p, b, m, ɸ, w]を唇音([labial])として一つの抽象的なカテゴリーとして扱っている可能性を示唆している．言い換えると、実在のオムツの名前に頻出する[p, m]に共通する素性として[labial]を抽出し、新しくオムツに名付けを行う際にも、この素性を適用させた可能性がある．つまり、個々の音から、弁別素性への一般化が行われた可能性があると言える (c.f., Albright 2009; Finley & Badecker 2009)．[ɸ, w]（特に前者）は、比較的習得が遅い子音であるにも関わらず (Ota 2015)、オムツの名前に適切とされたことから、音象徴は抽象的なレベルでのつながりも可能であることを示唆している．

実際、日本語音韻論の理論的な考察では、[w]が音韻素性として[labial]を持つかどうかという明らかな証拠は、今まで存在しなかった (Kumagai 2017)．日本語音声学・音韻論の入門書や概説論文において、[w]の調音点をどのように分類するかは異なっている (labial: Kubozono 2015; velar: Tsujimura 2014; labiovelar: Labrune 2012)．本実験の結果は、[w]が他の唇音と同じように振る舞うという点において、[labial]の素性を持つことを示唆している．（但し、Kumagai (2017)の連濁実験では、[w]が OCP-labial 制約の連濁阻止に関与しないことを実証しており、[w]は音韻論的に唇音[labial]ではないと考察されている．）今後、より多くの実験を重ねることにより、音象徴の振る舞いの分析から、日本語の/w/の音韻的素性に関する議論も可能になることが期待される．

5. 結論

本研究は、音象徴の観点から、赤ちゃん用オムツの名前について 2 つの実験を行い、いずれも「唇音＝赤ちゃん用オムツの名前」という音象徴的なつながりがあることを示した．また、本実験の結果は、日本語母語話者が、実在する赤ちゃん用オムツの例を基に、唇音([labial])という共通する素性を抽出し、新しいオムツの名付けに際して適用させた可能性を示唆している．この点において、本結果は音象徴的なつながりは弁別素性というある程度抽象的なレベルで起こることも可能であることを示唆している．

謝辞

本実験に協力してくれたすべての学生に感謝する．本研究は第二著者への JSPS Grant # 17K13448 への援助を受けて行っている．

参考文献

Albright, Adam (2009) Feature-based generalisation as a source of graident acceptability. *Phonology* 26, 9–41.

Baayen, R. H. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*.

Cambridge: Cambridge University Press.

Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler and Morten H. Christianson (2016) Sound-meaning association biases evidenced across thousands of languages. *PNAS*.

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christianson and Padraic Monaghan (2015) Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences* 19(10), 603–615.

Finley, Sara and William Badecker (2009) Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61, 423–437.

Hinton, Leane, Johanna Nichols, and John Ohala (2006) *Sound symbolism*. 2nd Edition. Cambridge: Cambridge University Press.

Jacobson, Roman (1941/1968) *Kindersprache, Aphasie und allegemeine Lautgesetze* (*Child language, aphasia and phonological universals*. The Hague: Mouton.)

川原繁人 (2015) 『音とことばのふしぎな世界』東京: 岩波書店.

川原繁人 (2017a) 「ドラゴンクエストの呪文における音象徴: 音声学の広がりを目指して」『音声研究』21(2).

川原繁人 (2017b) 『「あ」は「い」よりも大きい！？：音象徴で学ぶ音声学入門』東京: ひつじ書房.

Kawahara, Shigeto, Atsushi Noto, and Gakuji Kumagai (2016) Sound (symbolic) patterns in pokemon names: Focusing on voiced obstruents and mora counts. Ms. Submitted. Downloadable at http://ling.auf.net/lingbuzz/003196

Kawahara, Shigeto and Gakuji Kumagai (2017) Expressing evolution in pokemon names: Experimental explorations. Ms. Submitted. Downloadable at http://ling.auf.net/lingbuzz/003281

Kubozono, Haruo (2015) "Introduction to Japanese phonetics and phonology." In Haruo Kubozono (ed.) *The handbook of Japanese language and linguistics: Phonetics and phonology*. (pp. 1–40). Berlin: Mouton de Gruyter.

Kumagai, Gakuji (2017) Testing OCP-labial effect on Japanese rendaku. Ms. Revision submitted. Downloadable at http://ling.auf.net/lingbuzz/003290

Labrune, Lawrence (2012) *The phonology of Japanese*. Oxford: Oxford University Press.

Ota, Mitsuhiko (2015) "L1 phonology: phonological development." In Haruo Kubozono (ed.) *The handbook of Japanese language and linguistics: Phonetics and phonology*. (pp. 681–717). Berlin: Mouton de Gruyter.

Sapir, Edward (1929) A study in phonetic symbolism. *Journal of Experimental Psychology* 12, 225–239.

Tsujimura, Natsuko (2014) *An introduction to Japanese linguistics*. 2nd edition. Oxford: Blackwell.

B2

実際の発話者と知覚される個人性を切り分けられる音声の探索的検討

林　大輔（東京大学大学院人文社会系研究科）・山上　精次（専修大学人間科学部）
daisuke.semmu@gmail.com, yamagami@psy.senshu-u.ac.jp

1. はじめに

ヒトは音声を聞いた時に，何を話しているのかだけではなく，発話者が「どのような人であるのか」を知ることができる（e.g. Krauss, Freyberg & Moresella, 2002）．そのような発話者の個人性に関する情報は，大きく 2 つに分けて考えることができる．1 つは，年齢や性別などの「どのような発話者であるのか」という個人性一般に関する情報であり，もう 1 つは，「発話者が誰であるのか」という個人の同定に関わる情報である．これらの情報の処理は，相互に関わり合っていると考えるのが自然である．個人性に依存して発話者の声道などは生理学的に変化し，またその結果，音声の音響特徴も変化する．これらの変化は，個人性一般に関する情報でも，個人の同定に関する情報でも同様であり，両者は必ずしも明確に区別できるものではない．

しかし，神経心理学研究において，個人性一般に関する情報はある程度以上知覚でき，知らない発話者の音声を弁別することはできるにも関わらず，よく知っているはずの発話者の音声を聞いても同定ができない音声失認という症例が報告されている（e.g. Hailstone, Crutch, Vestergaard, Patterson & Warren, 2010; Roswandowitz et al., 2014）．これらの研究は，個人性一般に関する情報の知覚と，発話者の同定が異なる処理メカニズムによるものである可能性を示唆している．

そのため，音声から発話者の個人性に関する情報を知覚する過程を明らかにするには，これらの処理の独立性や階層性を調べることが不可欠である．しかしこれまでの研究では，「実際の発話者が誰であるのか」ということと，その人が発した音声から「どのような個人性が知覚されるのか」という関係が，ある程度 1 対 1 に対応するものとして扱われてきた．これ自体は，日常を考えれば自然なことではある．しかしそのような条件では，「どのような個人性であるのか」ということと「誰であるのか」ということが密接に繋がり合っているため，個人性一般に関する情報の知覚処理と，発話者の同定に関わる知覚処理をそれぞれ区別した上で検討するには不十分である．

そこで本研究では，同一の発話者であっても異なった個人性を持って知覚されるよう意図して発話された演技音声を用いて，「実際の発話者」と「知覚される個人性」を切り分けられる音声を見出すことを目的として，聴取実験を行った．そのような音声が得られれば，今後の研究において，音声から発話者の個人性に関する情報を知覚するメカニズムを明らかにする上で有用であると考えられる．実験では，地声で朗読された音声であれば，読まれている文章が異なっても似た個人性を持っていると評定されることを確かめた上で，同一の発話者であっても異なった個人性を持って知覚される音声が存在することを示す．

2. 方法

本研究では，2 つの実験を行った．基本的な方法は 2 つの実験で同一であったが，実験 1 では女性話者の音声のみを用いたのに対し，実験 2 では男性話者の音声も合わせて用いた点が大きく異なっていた．

2.1. 参加者

実験 1 には，26 名の大学生が参加した（男性 10 名，女性 16 名）．平均年齢は 19.58 歳で，標準偏差は 0.90 歳であった．また実験 2 には，実験 1 とは異なる大学生 26 名が参加した（男性 8 名，女性 18 名）．平均年齢は 21.08 歳で，標準偏差は 6.26 歳であった．全ての参加者が実験目的を知らなかった．また，全ての実験は専修大学人間科学部心理学科の基礎実験 2 の授業内に行い，参加者は事前に倫理に関する説明を受け，実験参加の同意書に署名を行った．

2.2. 装置

音声の加工と再生，および反応の取得に MacBookAir Early 2014 (Apple) を用いた．ヘッドフォン (Sony) を用いて音声を呈示し，液晶ディスプレイ (HP 2311) を用いて画面を呈示した．音声の加工は，Audacity (R) 2.1.0 と GNU Octave 4.0.3 を用いて行い，実験プログラムは jsPsych を用いて作成した（de Leeuw, 2015）．

2.3. 音声

7 人の話者について，3 種類ずつ（A，B，C）の音声を用いた（表 1）．

話者 1，2，および 7 の音声は，話速バリエーション型音声データベース（SRV-DB）内の，ATR25 文をプロが朗読した音声であった．話速は 8 モーラ/秒のものを用いた．話者 1 と 2 は女性であり，話者 7 は男性であった．A，B，C それぞれの音声は，別の文章を朗読したものであり，文章は話者間で共通であった．全て地声で朗読された音声であり，同じ発話者であれば似たような個人性を持って知覚されると考えられる音声であった．

話者 3，4，5，および 6 の音声は，同一の発話者であっても異なる個人性を持って知覚されるよう意図して発話された演技音声であった．具体的には，声優事務所である株式会社マウスプロモーションのボイスサンプルを，研究使用許諾を得て，実験者が適切と感じる音声を抽出して用いた．話者 3，4，5 は女性であり，実験者の聴取印象では，各話者の A の音声は幼い少女，B の音声は大人の女性，C の音声は少年の発話であるように知覚される音声であった．話者 6 は男性であり，実験者の聴取印象では，A および B の音声は青年，C の音声は少年の発話であるように知覚される音声であった．

このうち，実験 1 では話者 1〜5 の音声を，実験 2 では話者 4〜7 の音声を用いた．なおこれ以降，それぞれの音声は，たとえば話者 1 の音声を「音声 1A・音声 1B・音声 1C」といった形で表記する．

表1:各音声の情報の要約

話者	性別	データベース	実験者の聴取印象
1	女性	SRV-DB	全て地声による朗読
2			
3		マウスプロモーションのボイスサンプル	A(幼い少女) B(大人の女性) C(少年)
4			
5			
6	男性		A・B(青年)、C(少年)
7		SRV-DB	全て地声による朗読

表2:用いた表現語対

低い声	—	高い声
老けた感じの声	—	若い感じの声
女性的な声	—	男性的な声
張りのない声	—	張りのある声
弱々しい声	—	迫力のある声
細い声	—	太い声
澄んだ声	—	かすれた声
落ち着きのない声	—	落ち着きのある声

2.4. 音声の加工

全ての音声について，200ms 以上の無音部分は，200ms に短縮した．その上で，発話内容の影響を軽減するため，ランダムスプライシングを用いた（Scherer, 1971）．ランダムスプライシングは，音声の発話内容をマスクしながら，声質を保持する上で有用な音声の加工法だと考えられている（Teshigawara, 2004）．具体的には，音声を 250ms ずつのセグメントに分割し，各セグメントの立ち上がりと立ち下がりの 3ms ずつは，振幅を線形に変化させた．そのように加工した 250ms のセグメントを 20 個，もともと並んでいたセグメント同士が続くことがないように，ランダム順に並べて，5 秒間の音声を作成した．1 秒間の無音区間を挟んで，異なる順番で並べた 5 秒間の音声を繋げて，1 つの音声につき 11 秒間の加工音声を作成した．

2.5. 評定項目

木戸・粕谷（1999）に基づいて，声質を表現する日常表現語対を 8 つ用いた（表 2）．表現語対は「低い声–高い声」「老けた感じの声–若い感じの声」「女性的な声–男性的な声」「張りのない声–張りのある声」「弱々しい声–迫力のある声」「細い声–太い声」「澄んだ声–かすれた声」「落ち着きのない声–落ち着きのある声」であった．それぞれの表現語対について，「非常に・かなり・やや・普通・やや・かなり・非常に」の 7 件法を用いて，どちらの表現語対により近く知覚されるかの判断を求めた．評定値は，上記の内，左側の表現語が選択された時に小さく，右側の表現語が選択された時に大きくなるように 1〜7 で定量化した．

2.6. 手続き

実験は，1 人ずつ個室で行った．まず実験者が内容について教示を行い，参加者が課題を理解したら実験者は部屋の外に出て，参加者は 1 人で課題を行った．実験では，2.4 項のように加工した音声を 1 つずつ呈示し，それぞれの音声について，2.5 項で記述した 8 つの表現語対について評定を行った．1 つの画面に 8 つの表現語対を全て呈示して，参加者はそれぞれの表現語対について，マウスクリックを用いて 7 件法で評定を行った．8 つ全ての表現語対について評定を行った後，次の音声を呈示し，再び評定を行った．1 つの音声につき，1 人 1 回ずつ評定を行った．なお，音声の呈示順は参加者ごとにランダムであり，表現語対の呈示順や左右位置は音声ごとにランダムであった．

3. 結果

結果の解析の際に，ある音声に対するある参加者の評定値の一部に欠損値があった場合には，その参加者のその音声に対する評定値は全て削除した．欠損値を除くと，実験 1 は 1 つの音声当たり少なくとも 22 人分，実験 2 は 23 人分の評定値が得られた．各音声に対する各表現語対の評定値について，参加者間の平均をとり，得られた平均値を用いてまず，表現語を対象として平方ユークリッド距離についてウォード法を用いて階層的クラスター分析を行い，デンドログラムを作成した（図 1)．その結果，実験 1 と実験 2 で，類似した構造のデンドログラムが得られた．

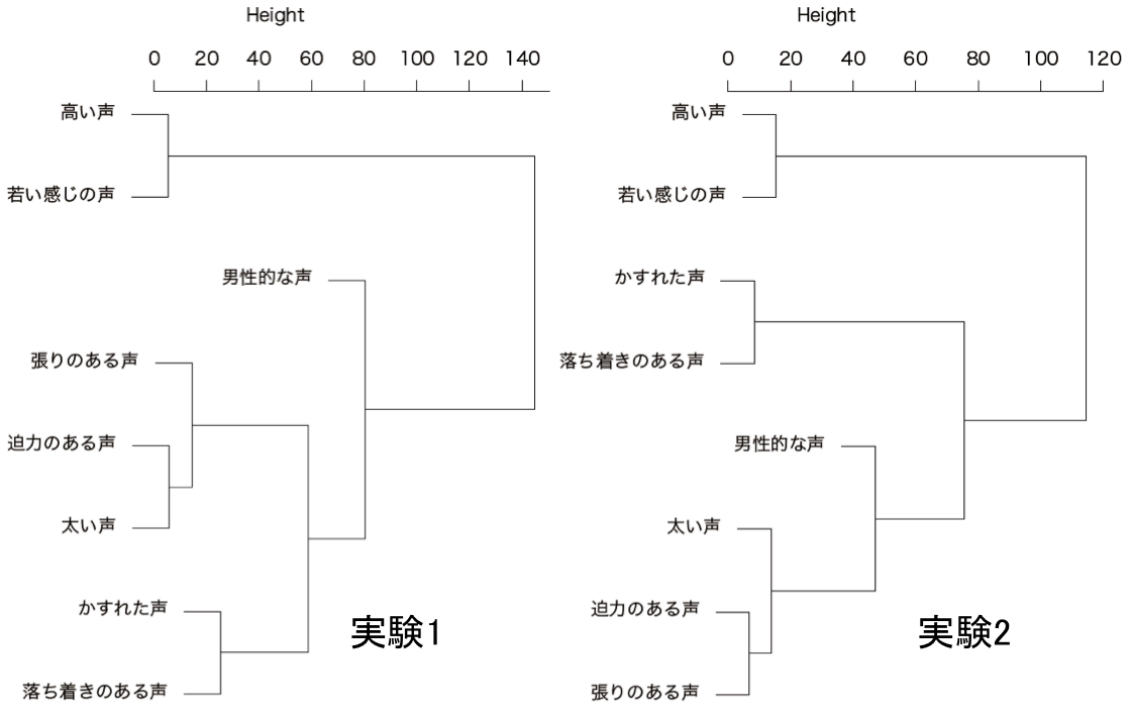


図 1: 表現語を対象としたクラスター分析の結果．(左)実験 1．(右)実験 2．

続けて，同様のデータを用いて，音声を対象としてクラスター分析を行い，デンドログラムを作成した (図 2 左，図 3 左)．得られたデンドログラムに基づいて，実験 1 では 3 つ，実験 2 では 4 つのクラスターに音声を分類した．実験 1 のクラスターはそれぞれ，実験者の聴取印象に基づくと，クラスター1 は少年，クラスター2 は幼い少女，クラスター3 は大人の女性の発話であるように知覚される音声が含まれていた．実験 2 のクラスターは，クラスター1 は幼い少女，クラスター2 は大人の女性，クラスター3 は少年，クラスター4 は青年の発話であるように知覚される音声が含まれていた．分類に基づいて，それぞれのクラスターに含まれる音声の，各表現語対に対する評定値を平均し，折れ線グラフで表した（図 2 右，図 3 右)．なお，折れ線グラフの背景は，図 1 に示したクラスター分析に基づいて，表現語対を 4 つのグループに分け，それぞれの区別がつくように灰色の濃さを変えた．

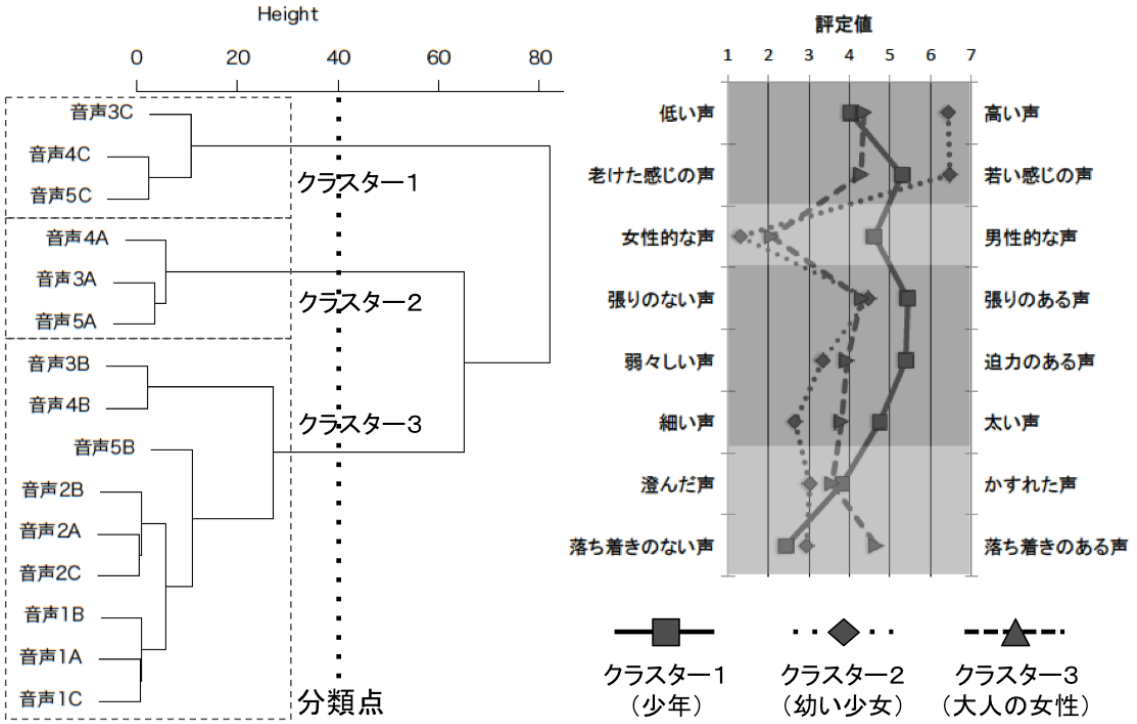


図 2: 実験１の音声の分類．（左）クラスター分析により得られたデンドログラム．（右）クラスターごとの評定値の平均．

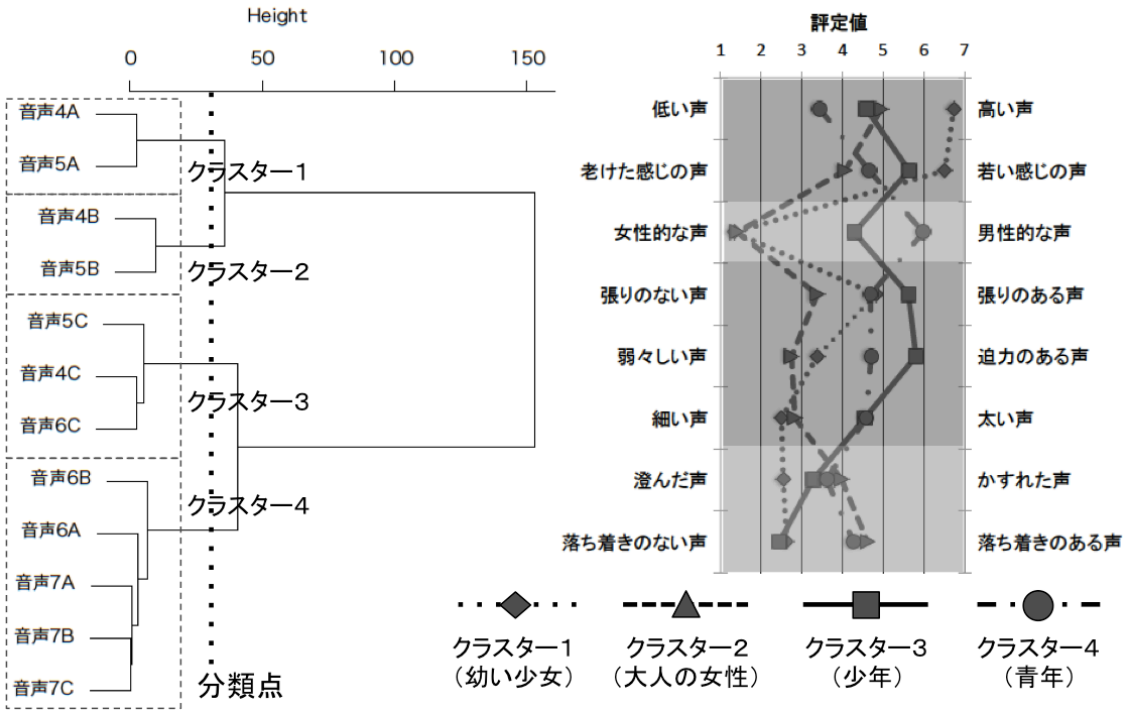


図 3: 実験２の音声の分類．（左）クラスター分析により得られたデンドログラム．（右）クラスターごとの評定値の平均．

4. おわりに

本研究の結果から，地声で朗読された音声は，同一の発話者であれば，文章が異なっても似た個人性を持って知覚されることが示された（図 2 左の音声 1A〜2C，図 3 左の音声 7A〜7C）．また，実験 1 では話者 3〜5 の 3 種類の音声がそれぞれ異なるクラスターに分類されたこと（図 2 左），実験 2 では話者 4〜6 の音声が異なるクラスターに分類されたことから（図 3 左），同一の発話者であっても異なった個人性を持って知覚される音声，すなわち「実際の発話者」と「知覚される個人性」を切り分けられる音声が見出されたと言える．

個人性の違いは声道などの生理学的特徴，ひいては音声の音響特徴に影響することから，音声の音響特徴と個人性の知覚との関係を調べることは，何が個人性の知覚に重要なのかを知る上で有用である（e.g. Kitamura & Akagi, 1995）．今回用いた音声の分類においては「性別」と「（声の高さと関連した）年齢」が強い影響を及ぼしていたことが見てとれるが（図 1 および図 2 右，図 3 右），音声の音響特徴について分析を行って知覚との関連を見ることで，何が個人性の知覚に重要なのかについて知見が得られると考えられる．

今回見出された音声を利用して，今後の研究において，たとえばある音声と発話者の結びつきを学習した時，知覚される個人性が異なっても同一の発話者であることが認識できるのか，などを調べていくことで，「実際の発話者」と「知覚される個人性」を切り分けた上で，音声から発話者に関する情報を知覚する過程を明らかにできるであろう．

参考文献

Hailstone, J. C., Crutch, S. J., Vestergaard, M. D., Patterson, R. D., & Warren, J. D. (2010). Progressive associative phonagnosia: A neuropsychological analysis. Neuropsychologia, 48, 1104-1114.

木戸 博・粕谷 英樹．(1999)．通常発話の声質に関連した日常表現語の抽出．日本音響学会誌，55，405-411.

Kitamura, T., & Akagi, M. (1995). Speaker individualities in speech spectral envelopes. Journal of the Acoustical Society of Japan (E), 16(5), 283-289.

Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. Journal of Experimental Social Psychology, 38, 618-625.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behavior Research Methods, 47(1), 1-12.

Roswandowitz, C., Mathias, S. R., Hintz, F., Kreitewolf, J., Schelinski, S., & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. Current Biology, 24, 2348-2353.

Scherer, K. R. (1971). Randomized splicing: A note on a simple technique for masking speech content. Journal of Experimental Research in Personality, 5(2), 155-159..

Teshigawara, M. (2004). Random splicing: A method of investigating the effects of voice quality on impression formation. Proceedings of Speech Prosody 2004, Nara, 209-212.

B3

日本のケーキ屋の売り子調の発声とその印象:
日本語母語話者と中国語母語話者の対照から

定延 利之（京都大学）・朱 春躍（神戸大学）・Donna Erickson (Haskins Laboratories, 金沢医科大学)・Kerrie Obert (Private Practice in Columbus, OH)
Sadanobu.toshiyuki.3x@kyoto-u.ac.jp; shu_s_y@koala.kobe-u.ac.jp, ericksondonna2000@gmail.com, kerriebobert@gmail.com

1. はじめに

我々の日常生活にはさまざまな発声法が観察される．それらの中には，他言語母語話者の耳には奇異に響くものもあるようである．或る音声が聞き手の母語によって異なる印象をもたらすとすれば，その原因は何なのだろうか？

この問題に関して，これまでによく知られているのは，意味論的な原因の存在である．たとえば Gumperz (1982)は，ヒースロー空港におけるインド・パキスタン系職員の解雇騒動を生んだ文化摩擦が，イギリス英語とインド英語における下降調イントネーションの意味の異なりに基づくと述べている．またたとえば，エリクソン・昇地 (2010) は，女性のアメリカ英語母語話者にとっての「感心」発話のイントネーション（焦点部分の冒頭が低く，その後で高くなる）が日本語の「疑い」発話と似ていると指摘している．

だが，こうした意味論的な原因とは別に，音声それ自体も問題の一因となり得るのではないだろうか？　本発表はこの可能性を追求しようとするものである．

具体的に取り上げるのは，日本の洋菓子店でケーキを売る若年層の女性話者が発する音声である．この音声を収録・観察した方法については次の第 2 節で紹介する．第 3 節では，この音声に対する日本語母語話者と中国語母語話者の印象が傾向として異なることを示す．その上で第 4 節では，この印象の違いが音声の多面性により説明可能であることを示す．

2. 音声収録・観察の方法

以下，刺激音声の作成，アンケート調査，音響分析の順に方法を述べる．日本語母語話者が抱く印象は既に調査済みなので(Sadanobu, Zhu, Erickson, and Obert 2016), 本発表では中国語母語話者が抱く印象と，日本語母語話者の印象との異同に焦点を置く．

2.1. 刺激音声の作成

刺激音声の話し手は 3 人の 20 代女性（大学学部生）の日本語母語話者（仮に P,・Q・R）である．このうち P と Q は実際に，関西の百貨店に入っている有名な洋菓子店でアルバイトしている同僚どうしで, 実験の時点 (P は 2013 年 11 月 22 日, Q は 2014 年 1 月 6 日）で, P は 3 年，Q は 2 年半の勤務歴があった．R にはケーキの売り子の経験は無いが，スーパーでレジ打ちをしていた経験が 8 ヶ月あった．P によると，ケーキを売る際の呼び声については，「明るく元気な声で」といった一般的なもの以外に特に指導は受けず，そこで P はアルバイトの先輩の発声を真似ていたと言う．（なお P は大学卒業後，その百貨店に就職した.）

録音した発話のセグメントは「いらっしゃいませ，どうぞご覧くださいませ」で，これ

を P・Q・R に営業用の声で(S)，そして「普段」の声で(U)，発してもらい，合計 6 種類の音声(PS・PU・QS・QU・RS・RU)を得た．

音声の収録は京都の ATR-BAIC で行われ，収録器として Marantz PMD 671，マイクとして Optoacoustics Optimic 1140 microphone が用いられた．

2.2. アンケート調査

母語話者の印象をとらえるために，以下 2 種類のアンケート調査がなされた：(1)ケーキ屋の売り子としてふさわしい度合いを 5 点満点（最低点が 1 点，最高点が 5 点）で問う調査 ; (2) 自由記述調査．日本語母語話者の場合，(1)は 110 人，(2)は（刺激音声によりばらつきはあるが）そのうち最多で 105 人，最少で 58 人の関西在住の大学学部生が回答した．中国語母語話者の場合，(1)は 38 人，(2)はそのうち最多で 35 人，最少で 28 人の北京在住の大学学部生が回答した．中国語母語話者は実験時点で日本語を習得中であった（初級 7 人・中級 15 人・上級 16 名）．

2.3. 音響分析

波形の音響分析には Wavesurfer software (Ver. 8.5.8)と SUGI Speech Analyzer（アニモ社）を用いた．.

3. アンケート調査の結果

アンケート調査に現れた，日本語母語話者と中国語母語話者の印象の異同を以下述べる．次の表 1 は，アンケート調査(1)の結果を平均点の形でまとめたものである．

表 1 によると，売り子のつもりで発せられた場合（つまり PS・QS・RS）の平均点の方が，「普段」の声で発せられた場合（つまり PU・QU・RU）の平均点よりも高いという点では両言語話者は共通している．

表 1: 6 つの刺激発話に対する日中母語話者の評価の平均点 （5 点満点）

評者＼発話	PS	PU	QS	QU	RS	RU
日本語話者	3.7	2.3	3.4	1.9	3.4	2.7
中国語語話者	2.3	1.8	3.8	2.3	3.0	2.3

だが，表 1 には日中両語話者の違いも現れている．ケーキ屋の売り子として，日本語母語話者の平均の評点は PS が最高だが，中国語話者の平均評点は QS や RS の方が高い．中国語話者による PS の平均評点（2.3 点）は，QU や RU の平均評点と同点である．この傾向は，中国語話者の日本語レベルにかかわらず一貫して観察される．

以上のような日中両語母語話者の類似と相違は，自由記述の調査(2)の結果を見ると，さらにはっきりする．日本語母語話者と中国語母語話者の自由記述の調査結果を，それぞれ図 1〜3・図 4〜6 に示す．

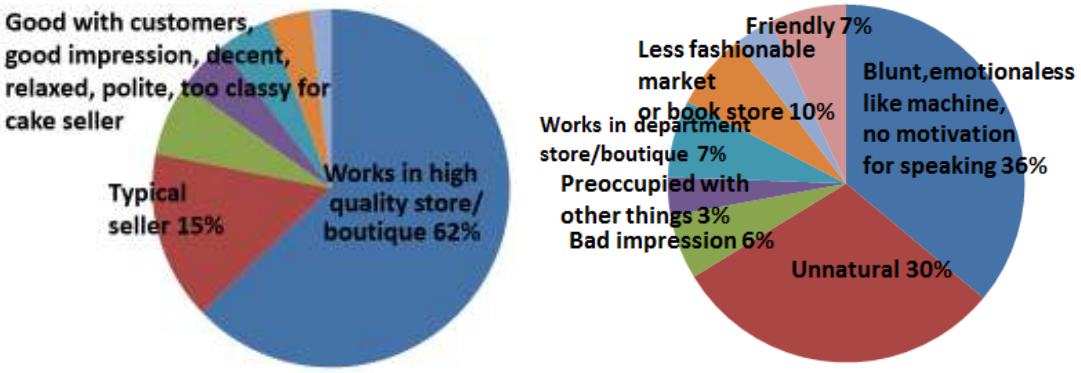


図 1:ケーキの売り子の声としての PS(左)・PU(右)に対する日本語母語話者の評価(自由記述).

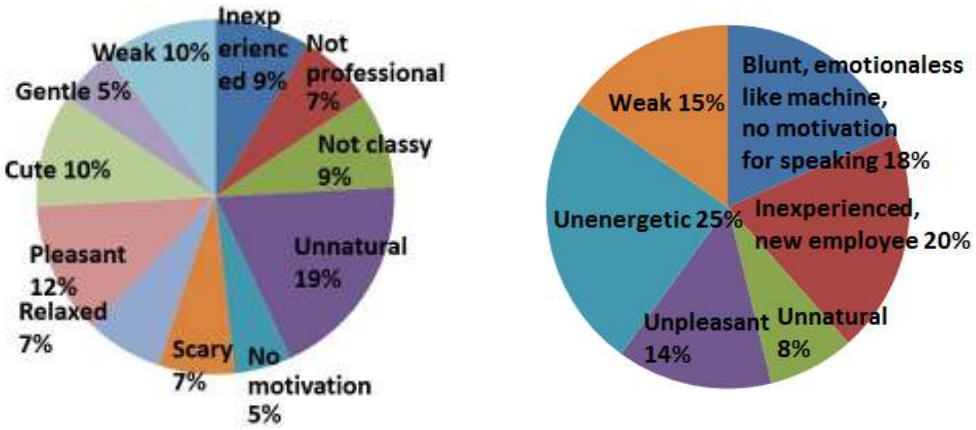


図 2:ケーキの売り子の声としての QS(左)・QU(右)に対する日本語母語話者の評価(自由記述).

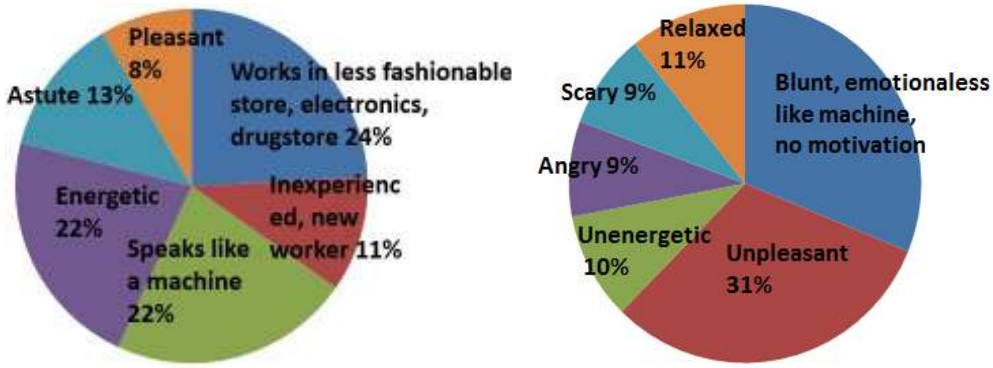


図 3:ケーキの売り子の声としての RS(左)・RU(右)に対する日本語母語話者の評価(自由記述).

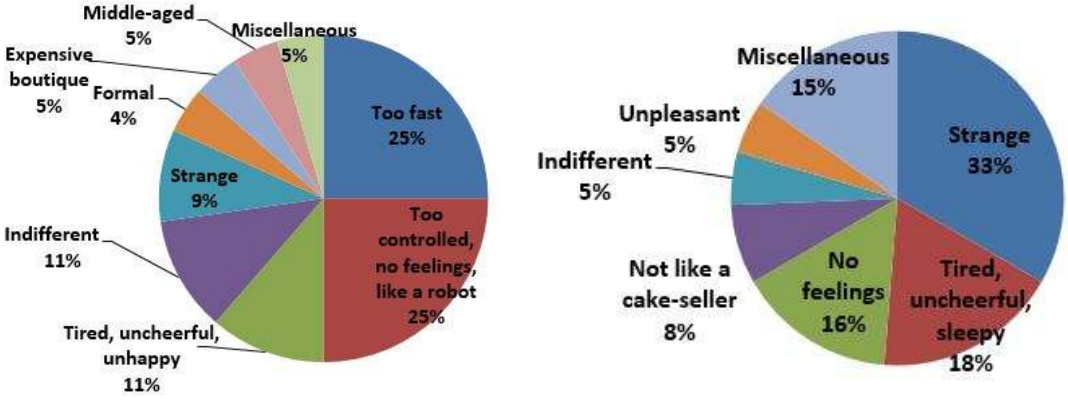


図 4:ケーキの売り子の声としての PS(左)・PU(右)に対する中国語母語話者の評価(自由記述).

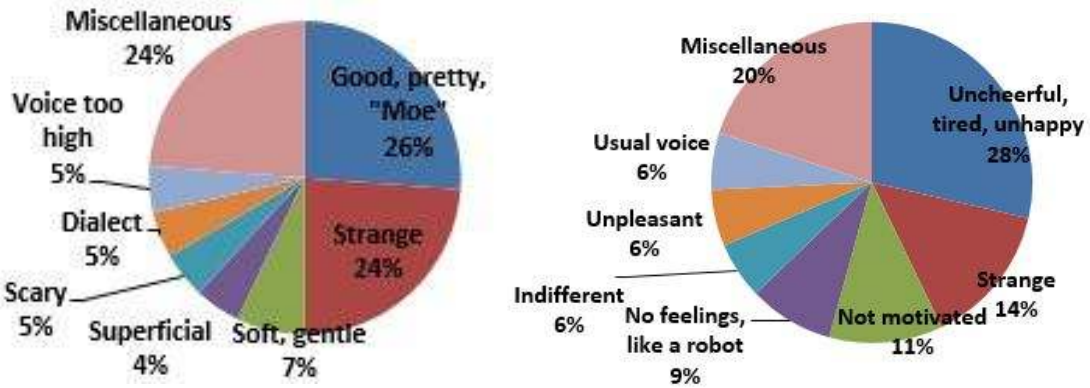


図 5:ケーキの売り子の声としての QS(左)・QU(右)に対する中国語母語話者の評価(自由記述).

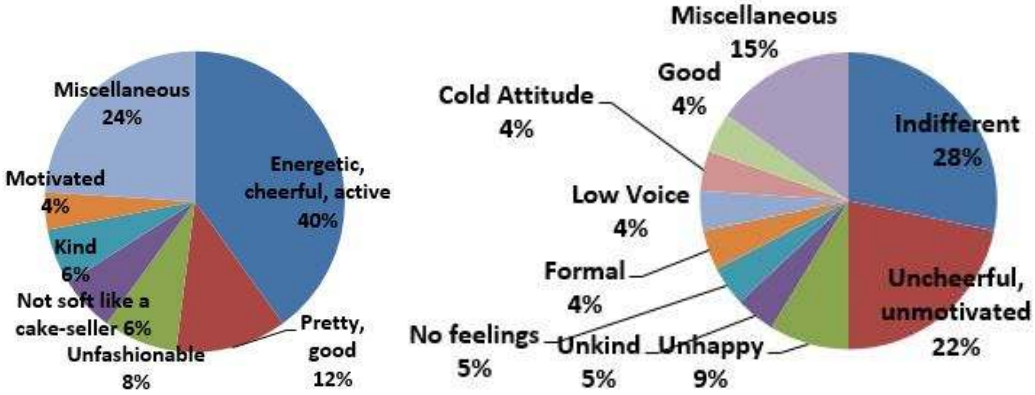


図 6:ケーキの売り子の声としての RS(左)・RU(右)に対する中国語母語話者の評価(自由記述).

左側（S）の円グラフの方が右側（U）よりも肯定的な記述が見られるという点では，日本語母語話者（図 1～3）・中国語母語話者（図 4～6）は一致するが，両者には違いもある．

図 1～3 の左側のグラフを見比べると，PS に対する日本語母語話者の評価（図 1 左）の高さが見てとれる．ここではケーキ屋の売り子らしさを認める記述が全体の 88%を占め（「ケ

ーキ屋の売り子らしい」「接客を心得ている」「いい印象」)，声の上品さ・楽しさ・落ち着き・丁寧さ・面白さを評価する記述が 10%あり，残りの 2%もこの声がケーキ屋にしては上品すぎるというもので，声自体の否定的な記述は無い．だが QS や RS はそうではない．

QS に対する自由記述は，「落ち着いている」「楽しげ」「かわいい」「やさしい」「かよわい」といった肯定に傾く記述も見られたが（45%），否定的な記述（「慣れていない」(9%),「プロっぽくない」(7%),「上品でない」(9%),「不自然な声」(19%),「やる気が無い」(5%), さらには「こわい」(7%)）が過半数を占めた（55%）．Q が実際にケーキ屋の売り子で働いているにもかかわらず，QS がケーキ屋の売り子の声らしくない，あるいはケーキ屋以前にそもそも売り子の声としてのふさわしさを疑う記述が見られた．

RS に対する自由記述にも，「売り子の声としてもより庶民的なスーパーや電化店，ドラッグショップの売り子」（24%）の他，「慣れていない」（11%）・「機械のよう」（22%）といった否定的な評価（33%）が含まれていた．RS に対する自由記述には肯定的なものもあるが，その中には「楽しげ」（8%）の他，「元気」（22%）・「活発/機敏」（13%）といった，品の良さとは必ずしも合わないものもある．スーパーのレジ打ちの経験しか無い R は，ケーキ屋とスーパーの違いをつけずに RS を発したのかもしれない．

業務用の声か「普段」の声かという違いに，日本語話者は敏感に反応したようである．全ての U 発話には，「無愛想」「機械のように感情が無い」「やる気が無い」（PU が 36%, QU%が 18%，RU が 31%），さらに「慣れていない」「不自然」「不機嫌」「元気が無い」「悪印象」「気もそぞろ」（PU が 40%，QU が 66%，RU が 41%）という記述が見られた．

P は「普段」の声が，デパートやブティックであれ（7%），スーパーや書店であれ（10%），従業員らしいと記述された唯一の話し手である．さらに「親しげ」（7%）という記述も見られる．それに対して QU は「弱々しい」（15%），RU は（「落ち着いている」（11%）という記述もあるが）「怒っている」「こわい」（17%）と評されている．

これに対して，図 4〜6 が示しているのは，中国語母語話者が特に PS と QS に関して日本語母語話者とは異なる印象を持つということである．

前述のとおり日本語母語話者は PS を肯定的に評価する傾向にあったが，中国語母語話者の少なくとも 81%は PS に対して否定的な評価をしている：「速すぎる」（25%）・「わざとらしい」「感情がなくロボットのよう」（25%）・「疲れていて暗い」（11%）・「無関心」（11%）・「不自然」（9%）．むしろ，中国語母語話者はケーキ屋の売り子として，PS よりも QS を評価しやすい．中国語母語話者が最高の評点を付けがちなのは QS である（表 1 の 3.8 点）．

4. 分析

日中両語母語話者による印象の違いが示しているのは，音声を多面的にとらえ，「着目されやすい面が母語により異なり得る」と考える必要性ではないか．こう考えれば両語母語話者の評価は次のように理解できる：日本語母語話者は声質に着目しやすく，ケーキ屋の売り子の声として”twang”な声 (Estill *et al*. 1983, Honda *et al*. 1995) が評価されやすい．他方，中国語母語話者は F0 の最高値に着目しやすい．各刺激発話の F0 の情報を表 2 にまとめる．

表 2: 6 つの刺激発話の F0 の最高値・最低値・差分・ピッチ幅.

F0 ＼ 発話	PS	PU	QS	QU	RS	RU
最高値 (Hz)	335	296	457	347	390	262
最低値 (Hz)	202	228	340	275	205	156
差分 (Hz)	145	65	117	72	185	106
ピッチ幅 (st)	8.76	4.52	5.12	4.03	11.13	8.98

PS は, F0 の最高値が QS（457Hz）・RS（390Hz）と比べて相対的に低く，ケーキ屋の売り子として否定的に評価されやすい．QS が「かわいい」「萌え」といった肯定的評価を受けやすいのは，Kawahara (2016) で述べられているように F0 最高値の高さが「萌え」の声の特徴であるためと考えられる．

日本には，売り物や売り場によって様々な売り子の声がある．それらの声は通文化的に比較検討してみる価値があるだろう．

謝辞

中国語母語話者へのアンケート調査に関して大工原勇人氏の多大な協力に感謝する．本発表は日本学術振興会の科学研究費補助金による基盤研究 ((A)15H02605，研究代表者：定延利之)，国立国語研究所の共同研究プロジェクト「対照言語学の観点から見た日本語の音声と文法」の成果を含んでいる．

参考文献

ドナ＝エリクソン・昇地崇明 (2010) 「パラ言語情報にみられる異文化間の知覚の相違」林博司・定延利之（編）『コミュニケーション，どうする？ どうなる？』pp. 138-153. 東京：ひつじ書房

Estill, Jo, Thomas Baer, Kiyoshi Honda, and Katherine Harris. (1983) “Supralaryngeal activity in a study of six voice qualities.” Proc. Stockholm Music Acoustics Conference, 157-174.

Gumperz, John. (1982) *Discourse Strategies*. Cambridge: Cambridge University Press.［ジョン＝ガンパーズ（著)，井上逸兵・出原健一・花崎美紀・荒木瑞夫・多々良直弘（2004 訳）『認知と相互行為の社会言語学―ディスコース・ストラテジー―』東京：松柏社］

Honda, Kiyoshi, Hiroyuki Hirai, Jo Estill, and Yoh’ichi Tohkura. (1995) “Contribution of vocal tract shape to voice quality: MRI data and articulatory modeling.” In O. Fujimura and M. Hirano (eds.), *Vocal fold physiology, Voice Quality Control*. (pp. 23-38). San Diego: Singular Publishing Group.

Kawahara, Shigeto. (2016) “The prosodic features of “tsun” and “moe” voices.” *Journal of the Phonetic Society of Japan* 20:2, 102-110.

Sadanobu, Toshiyuki, Chunyue Zhu, Donna Erickson, and Kerrie Obert. (2016) “Japanese “street seller's voice.”” The 5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Hilton Hawaiian Village Waikiki Beach Resort, Honolulu, Hawaii. Dec. 2. 2016.

B4

声道模型における共鳴の有限要素法解析

川原繁人（慶應義塾大学）・川原睦人（中央大学）・熊井 規 (RCCM)
kawahara@icl.keio.ac.jp

1. はじめに

本発表では、Chiba & Kajiyama (1948)により提唱され、荒井隆之氏（上智大学）が実際に作成した声道模型の共鳴を有限要素法によって解析する．この声道模型は、2016 年度の日本音声学会学術奨励賞を受賞したが、その理由の一つとして、音響音声学の入門教育にこの声道模型が非常に有効であることが挙げられる (Arai 2011 他)．荒井氏が作成した模型には様々なタイプが存在するが、一番シンプルな VTM-T20 型では、日本語の 5 つの母音が全て長方形の管の組み合わせによって作られている．音響音声学の基本の一つは管の中の共鳴を理解することにある．第一著者はこの模型を音声学入門の授業で積極的に用いている．その理由は、文系の学生相手の授業であっても、荒井氏の声道模型を使うと、管の中の共鳴を比較的簡単に理解できるからである．具体的には、まずは、「速度(c)＝周波数(f)× 波長(λ)」という直感的に理解しやすい関係式から始める．この関係式はf =「ある人の 1 秒あたりの歩数」、「λ＝歩幅」、「c=その人が歩く速度」という擬人化を用いると学生も理解しやすい．その関係式を理解させた後、$f=c/\lambda$ を導く．さらに三角関数の基礎を復習したあと、管の長さ L と λ の関係に関して、「管の片方が閉じていて、片方が開いている場合、$\lambda_n=(2n-1)L$ が成り立つ」ことを導く．この計算過程を難しいと感じる学生も少なくないが、荒井氏の声道模型を用いて、管を共鳴させ、その実際に共鳴した音の周波数を計算すると、上の計算式から導かれる値と一致するので、直感的な理解がなされやすい．特に声道内で狭めが起きない schwa のフォルマントは $\lambda_n=(2n-1)L$ の計算式のみから計算でき、また荒井氏の声道模型でも schwa を表現する菅が実装されているので、デモを簡単に行うことができる．schwa での計算を基礎に、二菅で表せられる「あ」の音響、三菅で表せられる他の母音の音響を順次教えることで、音響音声学の基礎を理解できる．

音声学を学ぶ（文系）学生への入門としては、この計算方法だけでも十分かもしれない．しかし、一方で母音の音響の工学的な解析として Arnela et al. (2016), Mancini et al. (2015)などによって行われた有限要素法を音響パターンに応用する研究がある．これらの先行研究を踏まえ、本稿では有限要素法を用い、荒井氏の声道模型における共鳴パターンを解析する．有限要素法では、領域をメッシュに分割し、各メッシュ間の関係を比較的簡単な関数で近似することにより微分方程式の近似解を得る (Kawahara 2016).

2. 方法

声道内の共鳴を解析するために有限要素法を用いた．この解析の基礎方程式は付録に示す Helmholtz 方程式である．有限要素は、図 1 に示したような 6 面体要素を用いた．この要素の補間関数は、α_1~ α_8 を未定定数として：

$$P = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 z + \alpha_5 xz + \alpha_6 yz + \alpha_7 zy + \alpha_8 xyz$$

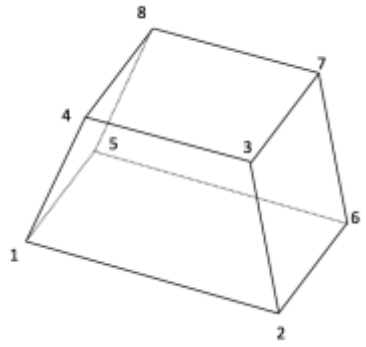


図 1：6 面体要素

である．有限要素法は、基礎方程式を重み付き残差方程式に変換し、補間関数を用いて、要素ごとの近似解を誘導し、これらを重ね合わせることにより解析する方法である．解析対象の形状を無理なく近似することができるため、多くの解析に用いられている．

反射境界と入射境界は問題なく処理できるが、Sommerfeld の無限遠における放射条件を処理するため、無限要素(infinite element)を用いた（この点において今回の解析は Arnela et al. 2016 や Mancini et al. 2015 と異なる）．解析の声道の形の寸法は、声道模型 VTM-T20 の設計図を使用した．メッシュは 6 面体要素（図 1）を約 30 万個個用い、出口での音の放射を再現するため、半球状の無限要素を用いた（図 2）．実際の計算は Actran (Release 17, Free Field Technologies)を用いた．

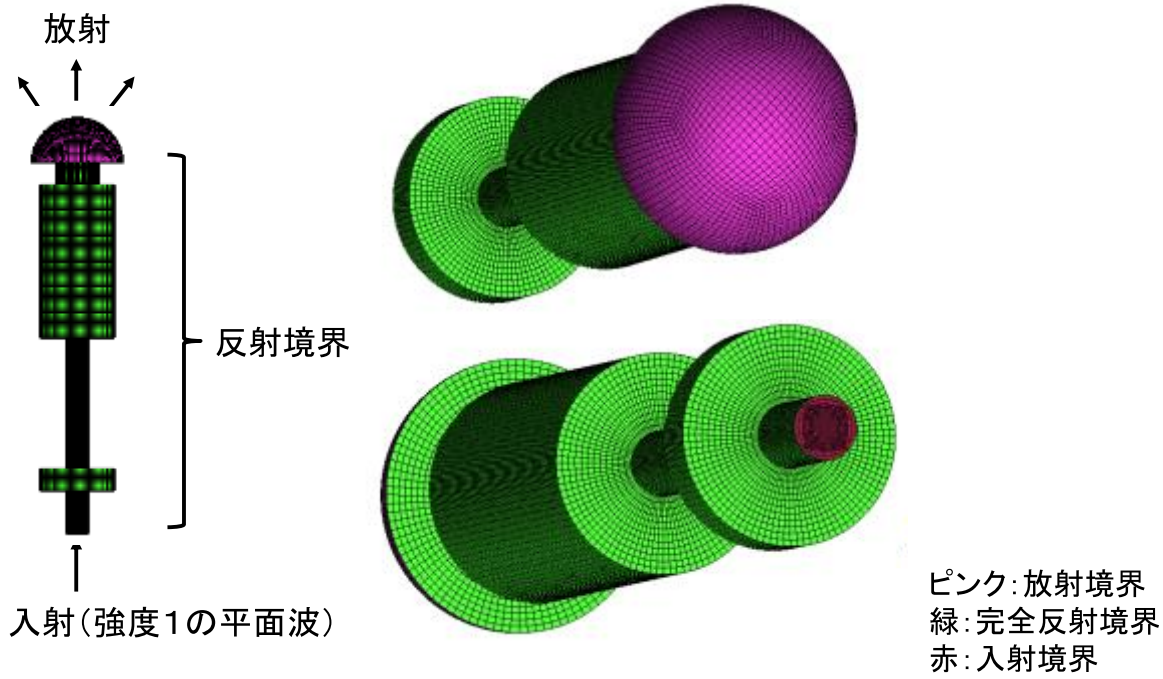


図 2：解析条件．「お」の管．

3. 結果

まず、日本語でもフォルマントの特徴が顕著に観察される「い」と「お」の結果について述べる．解析によって得られた図 3a の「お」のスペクトラを観察すると、図 4a,b の「い」よ

りも高い F1、そして後舌母音の特徴であるとても低い F2 が観察される．図 3a では、F1 と F2 がほぼ同じピークを成しているが、これは自然発話の/o/でも観察される現象である．例えば、図 3b に示した自然発話の「お」では F1 が 500 Hz、F2 が 700 Hz あたりに分布している．F3 の値も自然発話の「お」と非常によく合致しているのが分かる（2,500 Hz 付近）．

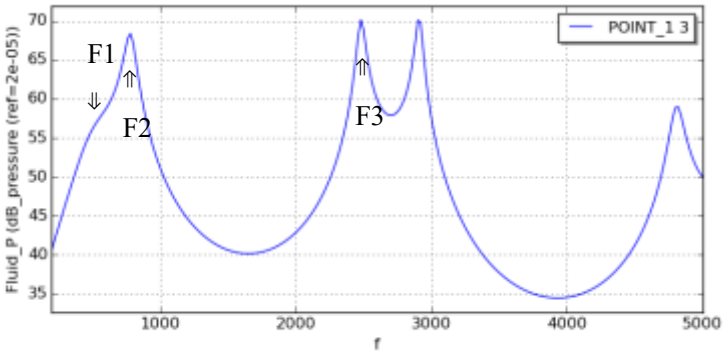


図 3(a)「お」の FEM 解析結果

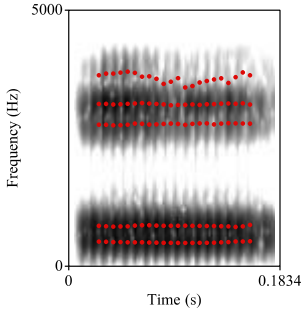


図 3(b) 自然発話の「お」

図 4a に示す「い」のスペクトラを観察すると、高母音の特徴である低い F1 と、前舌母音の特徴である高い F2 が観察されるのが分かる．ただし、図 4b の自然発話の「い」に比べると、F3 の値が若干低すぎる．

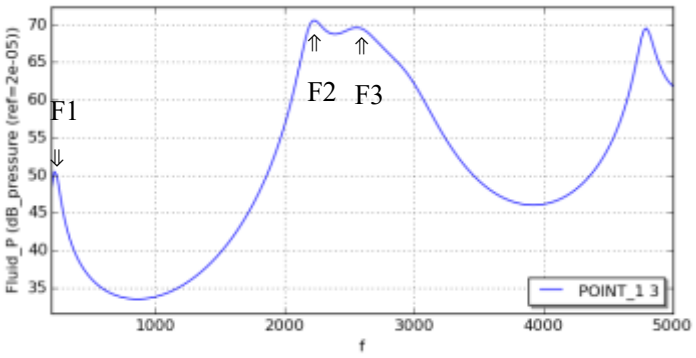


図 4(a)「い」の FEM 解析結果

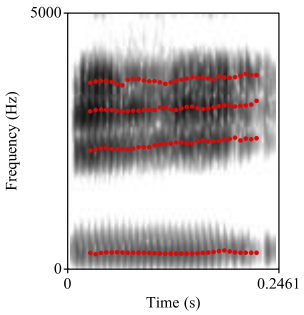


図 4(b) 自然発話の「い」

図 5 に「あ」の解析結果を示す．この解析結果では、「あ」特有の高い F1 はしっかり捉えられているものの、F2 が自然言語に見られる「あ」のそれよりもかなり高い（cf. 図 5(a) vs. 図 5(b)）．日本語の自然発話の「あ」では F2 が 1200−1500 Hz で観察されることが多いが、図 5(a)のスペクトルの二番目のピークは、2500 Hz に分布しており、これは、自然発話における「あ」の F3 を捉えてるようである．換言すると、今回の解析結果は「あ」の F2 をうまく再現していないが、「あ」の他の音響的特徴は捉えていると言っても良いかもしれない．

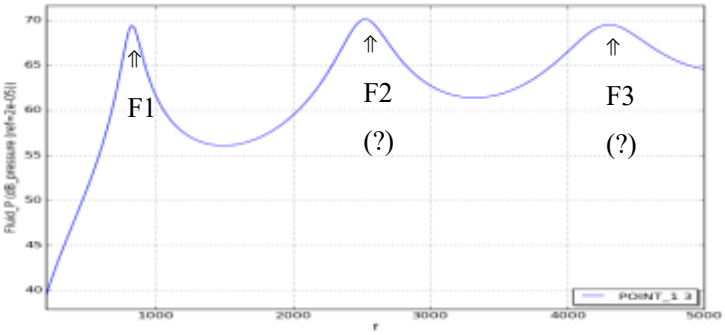


図 5(a) 「あ」の FEM 解析結果

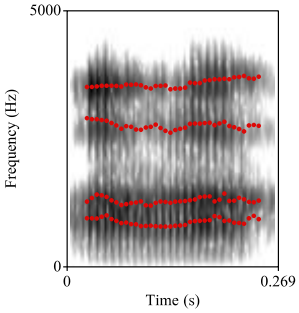


図 5(a) 自然発話の「あ」

4. 結論

本解析では、荒井氏作成の声道模型における圧力変化を有限要素法で 3 次元的に解析した．今回の解析では、「『い』の F3 が多少低く推定されてしまう」「『あ』の F2 をうまく再現できない」などの欠点もあったが、日本語の母音の基本的な特徴的を捉える第一歩としては、成功したと言えるであろう．「あ」の F2 をうまく再現できないという問題に関しては、今回使用した菅が全て長方形の形状をしていることに起因している可能性がある．すなわち、6 面体要素を用いて、管の角部に発生する渦流を解析するためには、より細かいメッシュを用いる必要があることが考えられる．また、無限要素と有限要部の接合部に生ずる誤差が影響している可能性もある．さらには、管の面と空気の摩擦も考慮に入れなければならない．よって現在、より丸みのおびた VTM-N20 の模型を使用し、再解析を行なっている．

まとめると、今回の解析では予備的調査として、VTM-T20 を用い、荒井氏の声道模型から発せられる音響特徴を有限要素法を用いて解析した．得られた音響のスペクトルと実際の音声を比較したところ、このような解析が有意義であることが示された．

参考文献

Arai, T. (2011) Education in acoustics and speech science using vocal-tract models. The Journal of the Acoustical Society of America 131(3): 2444-2454.

Arnela, M., Guasch, O., Dabbaghchian, S. & Engwall, O. (2016) Finite element generation of vowel sounds using dynamic complex three-dimensional vocal tracts. *23rd International Congress on Sound & Vibration*.

Chiba, T. & Kajiyama, M. (1941) *The Vowel: Its Nature and Structure*. Tokyo: Kaiseikan.

Kawahara, M. (2016) Finite Element Methods in Incompressible, Adiabatic, and Compressible Flows. Springer Tokyo.

Mancini, S., Astley, R. J., Gabard, G., Sinayoko, S., & Tournour, M. (2015) On the numerical accuracy of a combined FEM/radiating-surface approach for noise propagation in unbounded domains. *ICSV* 22.

付録：波動方程式の数学的基礎

空気中に、直行座標 0-xyz 系をとり、流速を u, v, ω、圧力を P、密度を ρ、時間を t とするとき、次の関係式が成り立つ。

運動方程式：

$$\begin{aligned} \rho\frac{\partial u}{\partial t}+\frac{\partial P}{\partial x}&=0\\ \rho\frac{\partial v}{\partial t}+\frac{\partial P}{\partial y}&=0\\ \rho\frac{\partial \omega}{\partial t}+\frac{\partial P}{\partial z}&=0 \end{aligned} \tag{1}$$

質量保存式:

$$\frac{\partial \rho}{\partial t}+\rho(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial \omega}{\partial z})=0 \tag{2}$$

状態方程式:

$$P=P(\rho) \tag{3}$$

音速定義式：

$$c^2=\frac{\partial P}{\partial \rho} \tag{4}$$

(3) と (4) から:

$$\frac{\partial P}{\partial t}=c^2\frac{\partial \rho}{\partial t} \tag{5}$$

(1) と (5) を (2) に代入して整理すると、以下の波動方程式を得る：

$$\frac{1}{c^2}\frac{\partial^2 P}{\partial t^2}-(\frac{\partial^2 P}{\partial x^2}+\frac{\partial^2 P}{\partial y^2}+\frac{\partial^2 P}{\partial z^2})=0 \tag{6}$$

A を定数として、圧力 P を次のように与える:

$$P = Asin(k_x x + k_y y + k_z z - \omega t) \tag{7}$$

式 (7) を (6) に代入することにより、次の関係が得られる：

$$\frac{\omega^2}{c^2} - (k_x^2 + k_y^2 + k_z^2) = 0 \tag{8}$$

$$\frac{\omega^2}{c^2} = k^2, \text{ where } k^2 = k_x^2 + k_y^2 + k_z^2 \tag{9}$$

式 (6) は (8), (9) より次のヘルムホルツ方程式となる:

$$\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + \frac{\partial^2 P}{\partial z^2} + k^2 P = 0 \tag{10}$$

境界条件は、壁の完全反射条件:

$$\frac{\partial P}{\partial n} = 0 \tag{11}$$

吹き口の圧力の固定条件：

$$p = p_0 \tag{12}$$

無限遠での Sommerfeld の条件：

$$\lim_{r \to \infty} [r(\frac{\partial P}{\partial r} - ikP)] = 0 \tag{13}$$

ここで i は虚数単位で $r = \sqrt{x^2 + y^2 + z^2}$ である。式 (10) を式 (11)-(13) の境界条件を満足するように解けば良い。このためには有限要素法を用いる。

B5

「私の日本語母音図」を作る

竹内 京子（國學院大學）
kyotake@kokugakuin.ac.jp

1. はじめに

音声学の授業を受講すると必ず学習する国際音声記号、特に母音図の役割は何であろうか？すべての母音の発音方法が１つの図にまとめられ、この図で世界中の言語の音が説明できるとされる。そのことに驚く学習者も多い。

教師が音声学の授業で発音記号を教える際、まず母音図を示し、どのように定義されているかを説明するのが一般的である。また、母音図を中心にして個人差が分布することも付け加えることが多い。しかしながら、その時、学習者はどのくらい自分の調音と同じだと考えているだろうか。音声学を学習していない学生に「他人に日本語５母音を発音してもらう方法を考えてもらう課題」を与えた結果を示した先行研究（竹内 2016）によると、予想外の様々な方法が観察された。学習者の実際の調音が母音図の誤差の範囲内ではない可能性も高い。また、母音図の定義では舌の山の位置だけが特に重要であるが、実際の発音では他の観点も必要かもしれない。

本発表では、音声学をまだ学んでいない被験者に日本語５母音の発音の際の舌の山の位置をプロットしてもらい、位置関係を調べた。

2. 先行研究

2.1. 自分の日本語母音の調音の推定（竹内 2016）

音声学や音響学の授業で母音図を学習する前の自分の調音を感じ、相手がその通りに行ったら同じ音が出るかどうかを確かめる実験である。「あいうえお発音大会」という形式で、グループ対抗でお菓子を使い、相手グループに自分が意図する日本語５母音を発音させる競争をした結果を示している。その結果、学生の実験のまとめの記述に以下のような音声学未修者に独特な調音のとらえ方が見られた。

・「あ」は口の奥が開いている
・「い」は「う」よりも歯が開いている
・「う」は唇を丸める
・「え」は舌を前に出す。舌先に注目
・「お」は唇を丸め、奥を広くする

また、使用したお菓子の使用法の例としては、以下のような例が多かった。

・「あ」：舌先を固定、奥が開くように大きな菓子を入れる
・「い」：細い菓子を歯で横にくわえる
・「う」：上下の歯をつけるために厚みのない菓子を歯でかむ

・「え」：舌を前に出すために上歯と舌で平らな菓子をはさむ
・「お」：奥を広くするために大きな菓子を奥に入れる、
　　または唇を丸めるための菓子をくわえる

特に、舌先や上下の歯の間隔に注目し、それによって口腔全体の形を変化させようとしている例が多かった。また、実際の口腔の空間のとらえ方も通常の母音図の舌の位置の例ではみられないような様々な例が見られた。

しかしながら、この実験後、音声学を学習した後の被験者に同じ「あいうえお発音大会」を行うと、これらの特殊な例がほぼ消え、音声学の知識に沿った例しかみられない傾向があった。

2.2. フランス語鼻母音の調音評価実験（竹内 2006）

音声学を学習していない被験者を対象にフランス語鼻母音の位置を母音図の簡単な説明後、母音図上で位置の推定をしてもらった実験である。まったくの音声学未修者であったにもかかわらず、母語でない言語の音声の母音図上の位置の推定がほぼ可能であることが示された。このことから、本当に相手の調音の様子を推定しているかどうかは分からないが、未知の言語の音色の弁別をするための道具としては母音図は非常に有効であることが示された。

3. 「私の母音図」の作成

これらの先行研究をふまえ、実際の学習者の日本語母音発音時の舌の山の位置の測定を行ったのが今回の「私の母音図」の実験である。

3.1. 被験者

音声学で母音図をまだ学習していない日本語母語話者の大学生 15 名、言語聴覚士養成の専門学校生 31 名を被験者とした。実験前に母音図の存在を簡単に説明し、この実験で舌の山の位置をプロットする理由を伝えた。

3.2. 実験方法

音声学または音響学の授業内で行った。被験者は２人ペアでお互いに測定者と被験者となった。被験者は壁に B５の測定結果を記録する紙を貼った横に立ち、日本語５母音を発音した。測定者はその時の舌の山の位置に棒つきキャンディの先端をあて、口腔の中の舌の山の位置を棒の先で平行移動したところに５つの点をプロットした。キャンディの棒は常に床と平行になるように注意し、舌の山の位置も測定者と被験者でお互いに確認し合意を得た場所にプロットした。それぞれの点が正確にプロットされるように、鼻と額の位置は５母音発音時に動かさないように注意した。

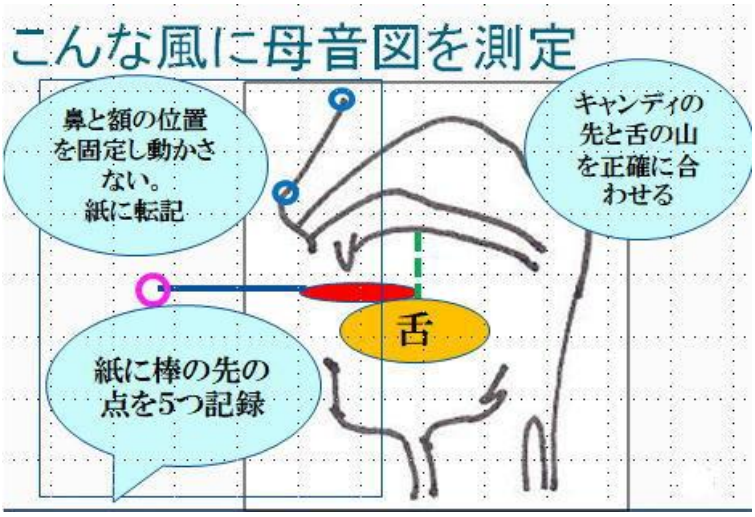


図 1: 母音図測定の様子

3.3. 結果の分析

個人ごとのデータを上下関係と前後関係に分け、５母音の順番を調べた。その後、一般に言われているものと違う点を探し、その数を集計した。

4. 実験結果

様々な５母音の位置関係のバリエーションがあった。図２が個人ごとのデータ、表１が特徴のまとめである。上下関係では、特に「い」の点が下の方にくるもの、「あ」が上にくるものが目立った。また、舌の山の前後では、前舌と後舌の母音の入れかえが頻繁に見られた。

狭広	K	N	合計
あ狭	7	8	15
い広	6	3	9
お狭	1	5	6
う広	2	2	4
え広	3	4	7
お広	3	0	3
い、えが逆	6	10	16
う、おが逆	6	7	13

前後			
う前	4	2	6
お前	4	2	6
い後	1	3	4
え後	3	3	6

表 1: 各母音の位置関係の結果のまとめ

K は大学生、N は専門学校生を示す。数値は図 2 における出現回数である。例えば、「あ狭」は「あ」の舌の山の位置が上がっていることを表し、「う前」は「う」の舌の山が前に移動していることを示している。

5. 考察

約３分の１の被験者が「あ」の舌の山が他の母音と比較して高いとし、約５分の１の被験者は「い」が他の母音より低いとしていた。他の母音が一般的に示されている母音図とは上下、左右、逆の位置にプロットされていた例も多かった。まず、考えられることは、母音図の横の線は、まっすぐに立った時に床と平行には位置していないのではないかということだ。もし、前母音側に前傾していれば多少は「い」の舌の山の位置が下がることも予想される。ただし、狭母音と広母音が逆転するほどに違うかどうかは疑問である。

次に考えられるのは、舌の山または全体の位置が本当に別の場所にあるという可能性である。本当は様々な調音方法が存在するのかもしれない。今回は被験者にお互いに測定をさせたので、誤差もあるだろうが、それにしても、これらの数値は無視できない値ではないだろうか。

母語を獲得すると、母語に関係のない音声の特徴を人は無視できるようになる。無視できるようになるから母語を獲得できるとも言える。外国語の場合も同じような現象が報告されている。母音図についても同じようなことは言えないだろうか。音声学を学んだことのない学生には見えることが、もしかしたら、すでに音声学を学んでしまった我々には見えなくなっているのかもしれない。だとしたら、もう一度、白紙の状態に戻りたいものである。

6. 今後の課題

今回の実験は授業内で行ったこともあり、全員が正確に測定できているかは断言できない。しかしながら、無視できないほどの例が示すように実際の発音は母音図に示すようには感じていない学生が多いことは確かである。詳細について別の方法で調べることが今後の課題である。今回のデータを使って別の方法で分析することも可能であろう。また、音声学で母音図を勉強している学生の結果と違うかどうかも調べてみたい。さらに、先行研究において、母音の弁別を母音図で測定でき、外国語の音声知覚のためのツールとしては優秀であることが示されている。自分の音声の生成と他人の音声の知覚の間にどんな違いがあるのかについても考えてみたい。

参考文献

竹内京子(2006)「日本人学習者のフランス語鼻母音調音の評価」,Etudes didactiques du FLE au Japon 2006 第 15 号 83-91.

竹内京子(2016) 「音声学未修者の日本語調音の推定」日本音響学会春季研究発表会講演論文集

図 2: 各被験者のデータ-1　　左が上下関係、右が前後関係

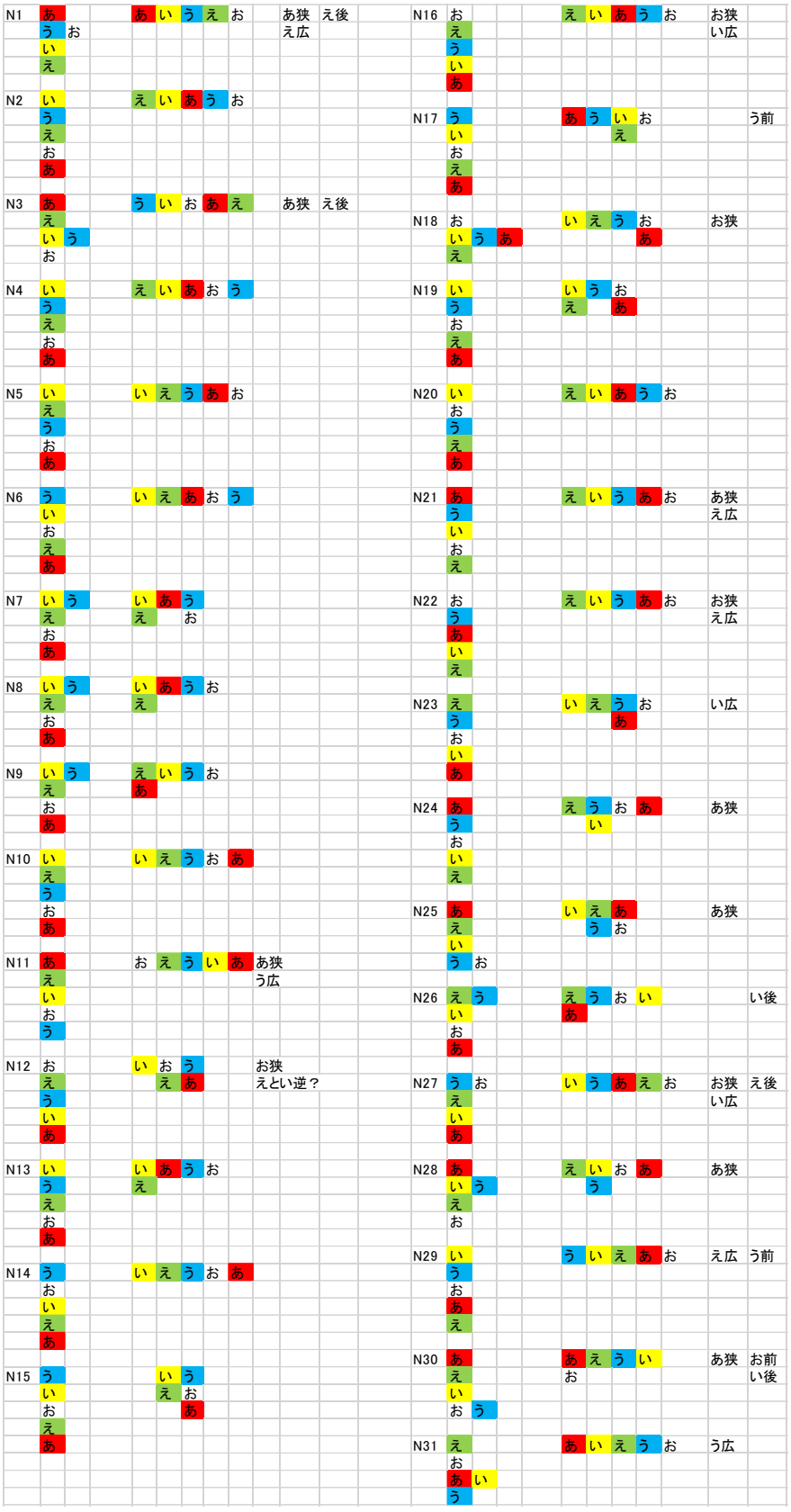


図 2: 各被験者のデータ-2

B6

F0 Contour Parameterization Using Optimal Regression Chains

Aaron Albin (Kobe University)
albin@people.kobe-u.ac.jp

1. Background

In a wide range of applications, both practical and scientific, it is useful to discretize an F0 track into a finite set of parameters that collectively represent a 'stylized' version of the shape of the F0 contour in the raw data. Such algorithms are used, for example, in machine-learning, automatic speech recognition, and computer-assisted language learning. The specific case examined in the present study is the automatic classification of a Japanese word (or phrase) based on its pitch-accent type. After first providing an overview of one common approach to this problem in Section 1, a novel approach that avoids several of its shortcomings is described in detail in Section 2. Section 3 then illustrates the method by applying it to a test dataset. Finally, in Section 4, the paper concludes with a discussion of promising directions for future development.

1.1. Mora-based F0 contour parametrization

One popular approach to the F0 contour parametrization problem involves first calculating some form of representative F0 for each mora (e.g., by averaging across the F0 points therein), then calculating the change in these values between each pair of adjacent moras - each resulting change value being one parameter. This approach is described, for example, in Ishi et al. (2003) as the "CV-average" operationalization of "F0mora". While this method has proven useful in a wide range of contexts, at least three problems can be identified. First, consonantal 'microprosody' often creates unreliable F0 information and may lead to distorted parameter values. Such cases are not uncommon for [s], where most frames therein are voiceless, and for voiced stops like [g], where most frames are often voiced but merely represent a perturbation. Second, the number of parameters extracted for a given word is relatively small, e.g. only two parameters for a three-mora word like *nimono* ('stew'). Representing all possible F0 contour shapes over the six segments in a word like *nimono* with only two parameters involves a significant loss of information. Third, a similarly nontrivial amount of information is lost by summarizing by representing each mora's F0 with a single average. This is most problematic in cases where important F0 changes occur inside a single mora, e.g. if the F0 rises during the onset consonant and then falls during the vowel.

1.2. Present study

The following section describes an alternative approach to F0 contour parametrization, called Optimal Regression Chains (or "ORC"), that overcomes these three problems. With the proposed approach, one parameter is calculated for each individual segment (rather than each mora). The goal of this method is to extract a set of parameters from the F0 contour of an utterance in a way that preserves the separation between phonologically distinctive categories (e.g., Japanese accent types).

2. Proposed method

The proposed algorithm is implemented as a function in the R programming language that takes four pieces of information as input for any given file: (1) the soundfile itself (in .wav format), (2) a file containing F0 information (e.g., a Praat Pitch object saved in plain text format), plus a matrix containing (3) segmentation boundaries (i.e., timestamps of the beginning/end of every segment) and (4) labels for each segmentation interval indicating which phone is contained therein.

2.1. Reliable vs. unreliable F0 information

The discussion above alluded to the fact that the F0 information for certain segments is inherently more reliable than for others. This information is built directly into the ORC algorithm by making a distinction between 'reliable' and 'unreliable' portions of the F0 contour. Doing so makes the modeling more conservative by avoiding using F0 information that is likely to be influenced by well-documented sources of noise. Recall from above that any token to be analyzed must first be parsed into labeled intervals. Each such interval is classified as [+/- reliable] by applying two 'checks' to its associated information.

The first check involves cross-referencing each segment label with two (non-overlapping) exhaustive lists of [+ reliable] and [- reliable] labels, both specified by hand. Note that the since the inventory of segments is language-specific, and conventions for segment labeling can be researcher-specific, this information should be prepared specifically for each individual analysis. In informal testing, the following lists were found to be effective at maximally separating segments in Japanese with reliable and unreliable F0:

Reliable: (1) Vowels like /a,i,u,e,o/, (2) Nasals like /m,n/, (3) Approximants like /w,j/.

Unreliable: (4) Voiceless obstruents like /p,t,k,ts,tʃ,ɸ,s,ʃ,h/, (5) Voiced obstruents like /b,d,g,dʒ,z/, (6) flap consonants like /ɾ/

The second check involves calculating the percentage of voiced frames in each interval and confirm whether it falls above some minimum threshold. Even vowels can occasionally lack robust F0 for a variety of reasons, e.g., phonological vowel devoicing in Japanese, or utterance-final creaky voice. For this reason, it may be wise to treat the F0 information as unreliable if too many frames have missing/NA F0 values. The threshold itself can be set to any arbitrary percent, but informal testing suggests a minimum threshold around 20% is effective. That is, if less than 20% of the frames within an interval are voiced, then the F0 information in that interval is treated as unreliable. (Note that this second check can be 'turned off' by simply setting the threshold to 0%.)

In order for a given interval to be treated as reliable, it needs to pass both of the above checks. In other words, it needs to have not only an appropriate label but also a sufficiently high percentage of voiced frames. Any interval failing either (or both) of these two criteria is treated as unreliable. In intervals thus determined to be unreliable, all F0 points are changed to N/A, i.e. making it identical to intervals containing 0% voiced frames in the raw data.

2.2. Creating line segments

Next, a line segment is created for every interval. The exact details of how the line segments are determined depends on how many F0 points fall into that interval (i.e., how many voiced frames there are). In the majority case where there are 2 or more F0 points in the interval, a linear regression is fit to these points (with x=time and y=F0). In the rare case that there are exactly 2 points, this regression is trivially identical to simply connecting those two points with a straight line. If there is only 1 point in the interval, a perfectly horizontal (zero-slope) line passing through that point is used – i.e., with a single F0 value held constant throughout the interval.

Line segments are created even for intervals with no F0 points - either from lacking F0 points in the raw F0 track or due to being NA-ed out for having unreliable F0 as discussed in Section 2.1. Since such intervals have no usable F0 information, F0 information from neighboring intervals is used to fill in the gap. If the interval in question is initial or final within the word/sentence token, the nearest regression endpoint is copied to fill in the missing F0 values (via 'constant extrapolation'). For example, in a word like *ki* 'tree', if the first interval (/k/) is missing F0 and the fitted regression line begins at 123 Hz in the second interval (/i/), then the beginning and ending F0 values for the first interval are set to 123 Hz as well. If there are multiple intervals with missing F0 (the first 3 segments in a token of *suki* 'like' with a devoiced /u/), constant extrapolation is applied across all of them.

Alternatively, if the interval without F0 points is medial within the token (i.e., anything but the first or last), the line segment is created by copying the values of the regression endpoints in the adjacent intervals. For example, in a word like *aki* 'autumn', with missing F0 for the /k/, if the regression line of the interval to the left (/a/) ended at 234.5 Hz, the beginning of the target interval (/k/) is assumed to be 234.5 Hz as well. If there are multiple medial intervals with missing F0, linear interpolation is used to fill in the gaps. For example, in a token of *deshita* 'was' with the entire sequence [ɕi̥t] deemed unreliable due to devoiced /i/, if /e/ ends at 190 Hz and /a/ begins at 100 Hz, then the line segments would be filled in as follows: /ʃ/=190-160 Hz, /i/=160-130 Hz, /t/=130-100 Hz.

2.3. Optimizing the junctions

The linear regressions described in Section 2.2 are fit on an interval-by-interval basis, in isolation of the F0 points in all other intervals. As such, the 'junctions' (i.e., points of union) between two neighboring intervals almost never match up. For instance, in the word *ao* 'blue', a linear regression fit to the F0 over the /a/ may end up at 140.1 Hz, and the regression over /o/ may begin at 149.9 Hz. The resulting model is physically unrealistic since, at the moment of the junction, it implies the speaker's pitch needs to be at two different values simultaneously. Moreover, by unnecessarily having two parameters at each junction rather than one, the model is arguably overfitting the data.

The proposed method overcomes this problem by using an optimization algorithm – the `optim()` function in R – to determine, for each junction, which exact F0 value would fit the raw data the best. The `optim()` function is set to `method="L-BFGS-B"` so that the parameter search is

'box-constrained', i.e. only certain ranges of values are considered. More specifically, for each junction, of the two competing regression endpoints, the smaller one (i.e. the one with the lower Hertz value) is rounded down, and this is used as the lower bound. Likewise, the higher one is rounded up, and this is used as the upper bound. For the above *ao* example, the lower value is 140.1, rounded down to 140, and the upper value is 149.9, rounded up to 150, hence the ultimately-chosen best-fitting parameter must be between 140 and 150 Hz. The values used to initialize the parameter search are the midpoints between each such pair of (unrounded) bounds, i.e., 145 Hz in this example. In this way, in the set of parameters used in `optim()`, there is one parameter for each junction, totaling to the number of segments/intervals minus one (e.g., 6-1=5 junctions for *nimono* 'stew'). Note that the very beginning of the utterance (more precisely, the beginning of the regression line inside the first interval containing F0 points, e.g., the beginning of /n/ in *nimono*), is not treated as a free parameter. Rather, this value is determined based on a linear regression constrained to pass through the [time,F0] point of the first junction (e.g., the junction between the first /n/ and the /i/ in *nimono*). The same is true of the last interval containing F0 points, whose regression endpoint is likewise determined based on a constrained regression that must pass through the last junction. (Note that the regressions run initially, as described in Section 2.2, are free of such a constraint.)

At each step in the search through the parameter space, the reconstructed model for the contour as a whole is created through linear interpolation between the F0 targets represented by the various parameters. Each resultant model (one for each step in the search) is then evaluated in terms of goodness-of-fit by calculating the median absolute deviation ("MAD") between the model F0 and the raw F0. Since the model contour is created through linear interpolation, which can generate an F0 point at any arbitrary point in time, the time sampling between the model F0 and the raw F0 is kept identical, making it possible to directly subtract one from the other. Since sign (positive vs. negative) is not important, the absolute value is then calculated for each deviation. In an effort to mimic perception, the resulting values are weighted so that high-intensity frames impact the resulting statistic more than low-intensity frames. (For further details on MAD, see Albin (2015, pp.83-84), which uses the same method.)

The overall end result of applying the ORC algorithm is the matrix with one row for each segment/interval and the following columns: (1) Label, (2) Reliable, (3) Time0, (4) Time1, (5) Hertz0, (6) Hertz1, (7) Cents, (8) Voicing, (9) Utilized. Column (1) contains the label for the interval in question, and (2) indicates whether that label is classified as reliable. Columns (3) through (6) contain the time and F0 ("Hertz") information for the beginning ("0") and end ("1") of the line segment inside that interval. Column (7) represents the change in F0 from Hertz0 to Hertz1 in cents (i.e. semitones ×100). Column (8) indicates the percentage of frames inside the interval that are voiced, from 0 to 1 (0% to 100% voiced). Column (9) indicates whether the F0 information inside the interval in question was utilized or not, based on the [+/-reliable] and voicing threshold criteria.

Of this rich information, at a bare minimum, only two pieces of information are necessary to represent a 'skeleton' of the entire contour shape: (A) the 'Hertz0' value of the very first segment, i.e. where the contour as a whole begins in Hertz space, and (B) the 'Cents' values for every segment in the contour. The former likely indexes things like speaker sex and emotional arousal, whereas the latter contains phonologically-relevant information about contour shape.

3. Application to test dataset

The materials used for the present test application were 160 tri-moraic Japanese nonwords, originally designed for another purpose. These words represented three different pitch-accent types: initial-accented (on the first mora), medial-accented (on the second mora), and unaccented. The 160 words were split into four subgroups of 40 words, each consisting of 20 minimal pairs: (1) unaccented-medial minimal pairs (*hetoya-he**tó**ya*), (2) unaccented-initial minimal pairs (*wakumi-**wá**kumi*), (3) initial-medial minimal pairs (*kózabe-koz**á**be*), and vocalic minimal pairs (20 pairs like *dohesa-dohosa*). Globally, the 160 words were roughly balanced for accent type: 52 were initial-accented, 52 were medial-accented, and 56 were unaccented. 140 words were CVCVCV, 14 words were VCVCV (e.g., *ateyu*), 4 words were CVVCV (e.g., *meobi*), and 2 words were CVCVV (e.g., *dotsua*). In terms of segmental makeup, across all 160 words, there were 480 vowels, 154 voiced obstruents (including the flap /r/), 162 voiceless obstruents, and 144 sonorant consonants. The approximately equal representation among the three different classes of consonants is crucial for illustrating the effectiveness of the proposed method. These 140 words were read aloud in a quiet room by three female native speakers of Japanese: two from Shizuoka prefecture and one from Hiroshima prefecture. Words were blocked so that everything within a block had the same accent pattern, thus making it easier to pronounce the non-words with the intended accentuation. In total, with tokens from 3 speakers for each of 160 words, 480 soundfiles in total were analyzed with the proposed method.

4. Results

Figure 1 is an example of what the output of ORC looks like, as applied to an initial-accented word (as evidenced by the peak at the end of the first [e]). The top panel is the waveform, the middle panel is the F0 track (where thicker, redder portions of the contour indicate higher-quality F0 information), and the bottom panel is the segmentation superimposed over the spectrogram. The solid black lines in the F0 track are those created through regression. Since there are no F0 points in the initial [z], the dotted grey line over [z] was created through extrapolation leftward from the beginning of the regression line for the first [e]. Likewise, the unreliable F0 information during the flap [ɾ] is ignored and instead the dotted grey line is filled based on the regressions in the surrounding vowels. Note that, on the whole, the straight-line model fits the raw data quite well, and the 'filled-in' values are plausible representations of what the missing F0 information might have looked like.

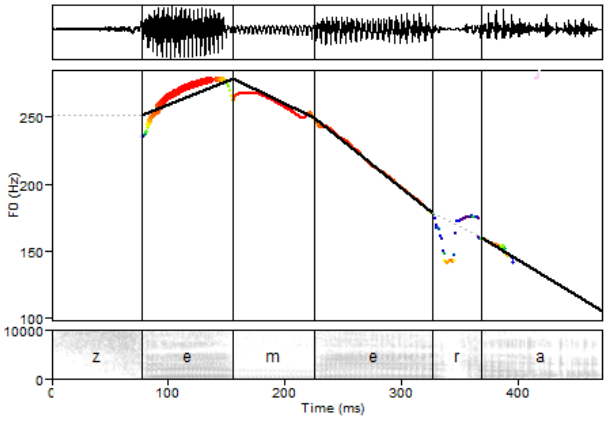


Figure 1: Example of output, applied to nonword ***zémera***

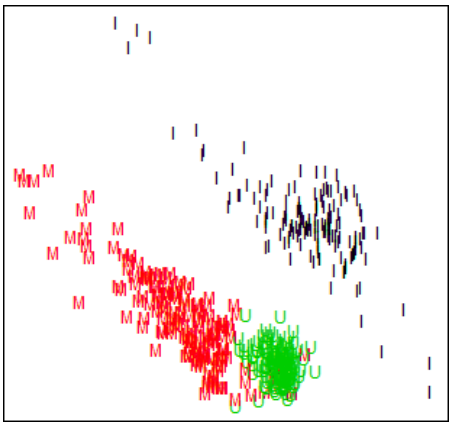


Figure 2: Multidimensional scaling results

The straight-line model in Figure 1 can be represented with the following seven parameters: 251, 0, 178.2, -192.1, -576.8, -188.9, -722. The first of these (251) is the initial F0 value (in Hertz), and the remaining six correspond to the size of F0 change over each of the six segments (in cents). Due to the makeup of the test dataset, these seven parameters can be estimated for nearly every token. (The only exception is the minority of VCVCV, CVVCV, and CVCVV words, for which the parameters for the missing consonants are N/A.) Setting aside the 'initial F0 value' parameter, the remaining six main parameters were visualized using multidimensional scaling, the output of which appears in Figure 2. There are 480 points in the plot – one for every token in the dataset. Initial-accented tokens are marked with black "I", medial-accented ones with red "M", and unaccented ones with green "U".

5. Conclusion

The clear separation between the 3 point clouds in Figure 2 is a testament to the effectiveness of the proposed method in extracting from the signal a set of parameters that maintains separation between the different phonological categories in question – thus attesting to how the proposed method successfully achieves its stated goal. This method holds much promise in a wide range of contexts, e.g., automated analysis of hard-to-classify productions by second language learners. Among numerous directions for future research, of particular importance is a systematic side-by-side comparison of the performance of the proposed method alongside other traditional methods.

References

Albin, A. (2015). Typologizing native language influence on intonation in a second language: Three transfer phenomena in Japanese EFL learners. Ph.D. dissertation. Indiana University.

Ishi, C. T., Hirose, K., and Minematsu, N. (2003) "Mora F0 representation for accent type identification in continuous speech and considerations on its relation with perceived pitch values", Speech Communication 41, 441-453.

C1

An Electroglottograph study of voiced plosives in two Japanese dialects

Haruya Ogawa, Michinori Suzuki, Haruka Takeyama(International Christian Univ.)
Seunghun J. Lee(International Christian Univ. and Univ. of Venda)

{c191332k,c191444k,c191476n,seunghun}@icu.ac.jp

1. Introduction

Voiced obstruents in northeastern Japanese are described as fully voiced compared to other varieties of Japanese (高田 2011). In this paper, we report the findings of a preliminary Electroglottograph (EGG) study that analyzes recordings of 4 speakers of two Japanese dialects: Kanto and Hokkaido. Acoustic studies of these dialects exist, but as far as we know there are no EGG studies that investigate the difference in voicing between these two dialects. Figure 1 shows EGG results of the token /higo/ 'sticks' produced by a Kanto speaker and a Hokkaido speaker. The vibration in the production of /g/ by the Kanto speaker shows that the vibration almost stops, whereas the vibration fully continues in the Hokkaido speaker's production of /g/.

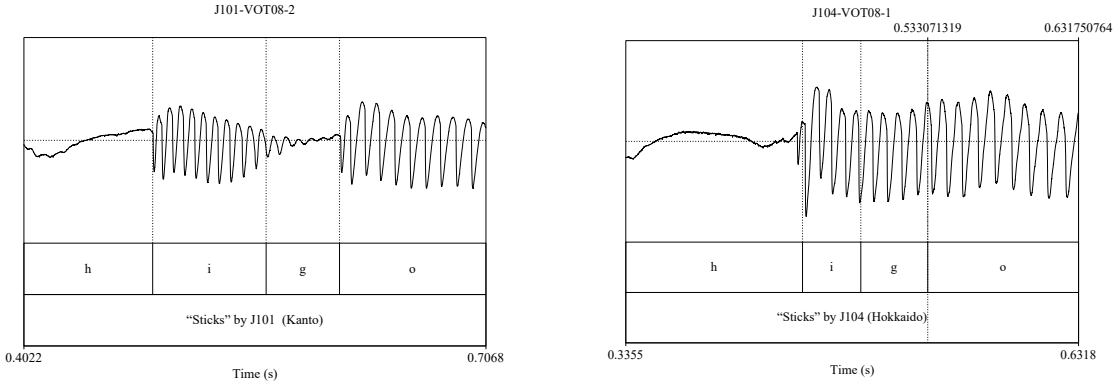


Figure 1: An EGG signal of /higo/ produced by a Kanto speaker (left) and a Hokkaido speaker (right).

Overall EGG results show that voiced plosives produced by Hokkaido speakers tend to have fuller voicing than those produced by Kanto speakers. While voiced plosives are fully voiced in Hokkaido speakers' speech, Kanto speakers show more variation: from no voicing to fully voiced. Examining the place of articulation, tone of a following vowel, and closure duration did not suggest any interaction with the voicing patterns.

2. Method

Two speakers from Hokkaido and two speakers from Kanto read 40 words that include three voiced

plosives [b], [d], [g] and three voiceless plosives [p], [t], [k]. All recordings were read three times in a frame sentence /kore-wa X desɯ/ 'This is X', where X is the target word. Stimuli written in Hiragana were presented in a slide show using power point, which was advanced by an experimenter. For the EGG recording, participants wore a two-headed electrode after applying electrogel on it. The electrode was connected to the EG2-PCX2 Electroglottograph machine, which was connected to an NEC Windows laptop computer and an ECM 8000 Microphone. Recording was done using Praat (Boersma 2001) on a laptop computer while the microphone was held by another experimenter. The analysis comes from a data set of 200 voiced plosives and 71 voiceless plosives. The voicing patterns of the plosives were analyzed using the EGG signal. Based on the voicing patterns, a number was assigned: from 1 for full continuous vibration to 0 for no vibration. Further categorical vibration patterns were analyzed as 0.2, 0.5 and 0.8 depending on the duration of the vibration within the closure of the plosives.

3. Results

3.1. Voicing

We first investigated the degree of voicing in Hokkaido speech versus Kanto speech. A cumulative link mixed model was used for analyzing the voicing pattern, with speaker as a random factor. Hokkaido speakers produced more tokens with full voicing than Kanto speakers ($p < 0.001$). This result holds regardless of whether all consonants are considered or whether only the voiced plosives [b], [d], [g] are considered. Table 1 shows the distribution of vibration pattern based on each plosive, only tokens with full vibration (1) and no vibration (0) are listed. Voiceless plosives show no vibration of the vocal cords in both dialects. In the Hokkaido dialect, almost all voiced plosives show full vibration in the EGG signal. In Kanto, [b] is always fully vibrating, whereas EGG data of [d] and [g] indicate that non-labial voiced plosives may not always be accompanied with vibration of the vocal folds. Figure 2 shows this diribution in a histogram format.

Table 1: Tokens of voiced and voiceless plosives.

	Hokkaido		Kanto	
	full vibration	no vibration	full vibration	no vibration
[b]	60	2	54	0
[d]	47	0	33	3
[g]	82	0	57	13
[p]	0	4	0	5
[t]	2	3	0	6
[k]	0	60	0	59

3.2. Voicing and place of articulation

We then looked into the effect of place of articulation on vibration. In the Kanto dialect, the distribution in table 1 shows that the voiced velar plosive has more tokens with no vibration, compared to the voiced bilabial plosive where there is no tokens with no vibration. In the Hokkaido dialect, nearly all voiced plosives show full vibration, but two instances of no vibration were observed in the bilabial [b]. The cumulative link mixed model shows that there is no overall effect of place of articulation, nor is there an interaction between voicing and place of articulation in both dialects.

3.3. Voicing and lexical tone

Both the Hokkaido dialect and the Kanto dialect have a lexical pitch accent. We examined whether there is an effect of lexical tone in the following vowel in the degree of vibration in the Kanto dialect. The model shows that there is no overall effect of tone in predicting the vibration in either dialect. However, there was an interaction found between dialect and lexical tone. The difference in the vibration of voiced plosive before low versus high tone in the Kanto dialect is larger than that in the Hokkaido dialect. This effect, nonetheless, is marginal ($p = 0.04$)

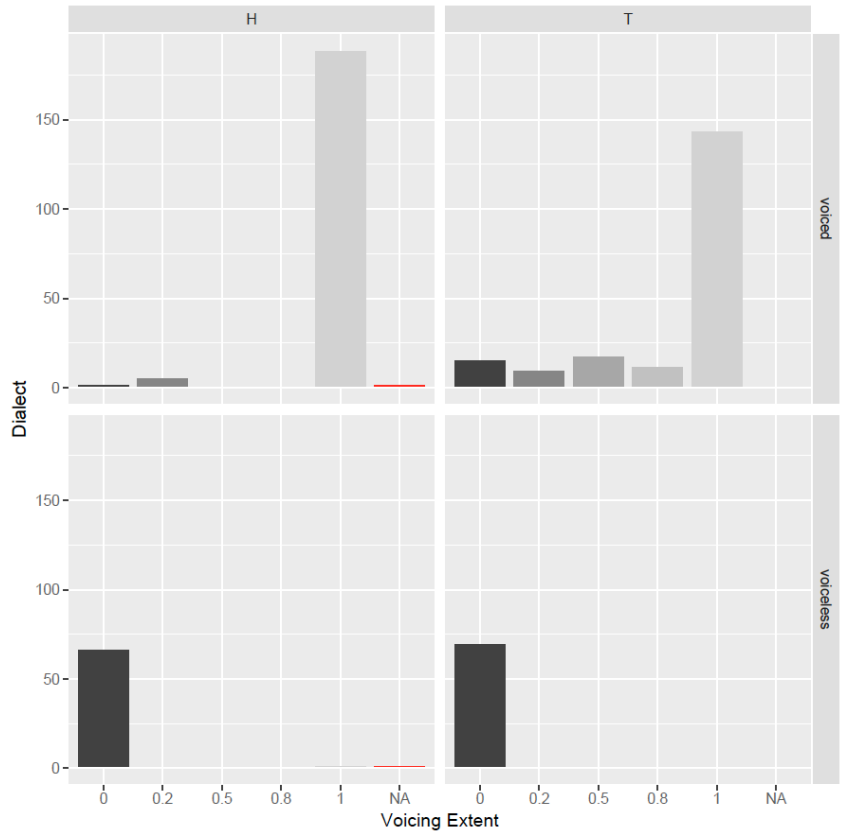


Figure 2: Histogram of the distribution of vibration in voiced versus voiceless plosives in Hokkaido. The y-axis shows the number of tokens belong to a category: 1 for full vibration and 0 for no vibration.

3.4. Voicing and closure duration

Voiceless plosives and voiced plosives have different closure durations. An analysis of the three voiced plosives [b], [d], [g] and three voiceless plosives [p], [t], [k] shows that the closure duration is significantly longer by about 45 ms across both dialects for voiceless stops ($p < 0.001$). We hypothesized that the closure duration may compensate for some neutralized vibration in the Kanto dialect. We tested whether there is an interaction between the degree of vibration and closure duration in both dialects using a standard linear mixed model with speaker as a random factor. The model suggests that an interaction exists for vibration and dialects, but in an unexpected direction: the closure duration difference for voiced versus voiceless plosives is slightly larger in the Hokkaido dialect than in the Kanto dialect, but only by 7 ms ($p = 0.02$). This means that the difference in the closure duration does not compensate for some neutralized vibration in the Kanto dialect.

4. Conclusion

This paper has shown the voiced plosives in the Kanto dialect show more variation when the vibration of the vocal cords is concerned. About 11% of the voiced plosives show no vibration, and most of them are velar plosives. In the Hokkaido dialect, however, no velar plosive shows no vibration, and only two labial plosives show no vibration. The results of a mixed model analysis show that there are no effects of place of articulation, a marginal effect of lexical tone, and a weak effect of closure duration.

Appendix. List of stimuli

HL (n = 21)			LH (n = 19)		
だれ	ぶな	まど	むだ	まご	かべ
ひご	ばね	がす	びら	げぷ	ぼく
しが	げき	かじ	ごみ	だめ	ひげ
べる	ぐみ	もず	かぎ	みず	どく
じが	ふぐ	ねじ	にじ	ばら	かび
はだ	そば	まだ	がき	さが	
へび	とび	どき	じく	ぐち	

Acknowledgement

We thank Jeremy Perkins for running the statistical analysis. This research is supported by JSPS Kakenhi 16K02641 and AKS-2016-LAB2250004.

References

高田三枝子. (2011) 日本語の語頭閉鎖音の研究 - VOT の共時的分布と通時的変化. くろしお出版.

Boersma, P. (2001) “Praat, a system for doing phonetics by computer.” Glot International 5: 341-345.

C2

熊本方言話者における促音音声詳細の世代差

高田 三枝子（愛知学院大学）takadam@dpc.agu.ac.jp

1. 研究の背景と目的

本稿では熊本の話者による促音＋破裂音という音連続における閉鎖区間の有声性に関わる音声詳細について話者の属性、特に世代差に注目し報告するものである。

日本語の首都圏方言を母体とする共通語においては、促音の生起環境に制約があり、/k, t, p/しか後続を許されないことが知られるが、一方で特に九州を中心とする西日本の方言でこの制約に反する音連続が方言語彙の中に見られる方言もあることが知られる（高山倫明 2012 など）。高山（2012）はこの音素配列上の制約が古くは後続音の鼻音性に対するものであったと仮定した上で、濁音の歴史的な音声変化の結果、制限に対する解釈および音声的な在り方が地域によって異なったという解釈を提案し、また先行研究に見える濁音の後続する促音（以下、有声促音とする）の分布を整理し地図化して示している。本研究で扱う熊本方言は、高山の説によれば、濁音の鼻音性の消失により促音の鼻音性への制限に抵触しなくなり、これによって有声促音の生起に問題がなくなった方言の一つであるとされ、高山の示す有声促音の分布図においてはこれを比較的多く持つ方言であることが見える。

一方、有声促音の音声詳細を音響分析の手法で観察し地理的な差異について指摘したものとしては、松浦年男（2016）、高田三枝子（2014、印刷中）などがあるが、ここでは紙幅の制限上、高田三枝子（2014、印刷中）のみ取り上げ紹介する。高田（2014、印刷中）は促音を含む外来語の発話について東北、関東、近畿、九州の話者それぞれ 13～14 名の音声を分析し、促音および後続する破裂音の閉鎖区間の音声詳細について地域差が存在することを指摘した。すなわち、まず有声促音において閉鎖区間中 voicing のパターンの出現頻度に地域差が見られ、voicing が閉鎖区間中途切れず続くパターン（full voicing: FV）は秋田＜東京＜大阪＜熊本と西の地域ほど多く発話されていた。また大阪の話者では、後続子音の prevoicing とみなせる voicing を伴うパタンが他地域より多く見られた。一方先行母音からの持続的な voicing の長さに関しても地域差が見られ、秋田＜東京＜大阪＜熊本と西の地域ほど長かった。さらに閉鎖区間長率（先行母音＋促音に対する比率）の有声／無声促音間の差が、やはり秋田＜東京＜大阪＜熊本と、西の地域ほど大きくなっていた。

以上のように、地域差については報告されているものの、各地域における世代差についてはその有無および詳細がいまだ不明である。世代差は通時的変化を反映している可能性もあり興味深い問題である。そこで本研究では、手始めに、広い世代にわたる熊本方言話者の促音＋破裂音（有声・無声のいずれも含め）の音声を分析し、その世代差について報告する。

2. 資料と分析

資料は 2006～7 年にかけて収集した録音資料「指標地域録音資料」（詳しくは高田（2011）参照）の一部で、今回分析するのはそのうち熊本県熊本市および八代市を中心とする北部方

言の 55 名の話者の発話資料である。全話者とも県外への移住歴はない。次の表 1 に話者情報を示す。資料には年代について偏りがあり、特に 70 代以上の話者は少ないため、分析結果の解釈時には注意を要する。

表 1 分析対象話者

話者記号	性別	生年	年齢 (調査時)	出身 市町村
KYS16M1	男	1990	16	水俣市
KYS16M2	男	1989	16	八代市
KYS16M3	男	1989	16	熊本市
KYS16M4	男	1989	16	八代市
KYS21M1	男	1985	21	八代市
KYS21M2	男	1985	21	八代市
KYS21M3	男	1985	21	八代市
KYS21M4	男	1985	21	水俣市
KYS21M5	男	1984	21	八代市
KYS29M	男	1976	29	八代市
KYS30M	男	1975	30	熊本市
KYS32M	男	1974	32	熊本市
KYS36M	男	1970	36	熊本市
KYS37M	男	1968	37	八代市
KYS39M	男	1967	39	熊本市
KYS41M	男	1965	41	八代市
KYS43M	男	1963	43	八代市
KYS45M	男	1961	45	八代市
KYS47M	男	1958	47	熊本市
KYS54M	男	1952	54	八代市
KYS56M	男	1949	56	熊本市
KYS58M	男	1948	58	八代市
KYS59M	男	1947	59	八代市
KYS67M1	男	1939	67	熊本市
KYS67M2	男	1939	67	八代市
KYS68M	男	1938	68	八代市
KYS79M	男	1926	79	八代市

話者記号	性別	生年	年齢 (調査時)	出身 市町村
KYS16F	女	1990	16	八代市
KYS19F	女	1986	19	熊本市
KYS21F1	女	1984	21	八代市
KYS21F2	女	1984	21	熊本市
KYS22F1	女	1984	22	宇土市
KYS22F2	女	1983	22	八代市
KYS22F3	女	1983	22	人吉市・八代市・牛深市
KYS25F	女	1980	25	八代市
KYS33F	女	1973	33	八代市
KYS34F	女	1971	34	八代市
KYS35F	女	1970	35	熊本市
KYS37F	女	1969	37	熊本市
KYS45F	女	1961	45	八代市
KYS49F	女	1957	49	八代市
KYS54F	女	1952	54	八代市
KYS56F	女	1950	56	八代市
KYS57F1	女	1949	57	八代市
KYS57F2	女	1949	57	八代市
KYS58F	女	1948	58	熊本市
KYS59F	女	1947	59	八代市
KYS60F	女	1946	60	八代市
KYS61F	女	1944	61	熊本市
KYS63F	女	1942	63	熊本市
KYS76F	女	1930	76	八代市
KYS78F	女	1927	78	八代市
KYS84F	女	1922	84	八代市

分析対象語彙項目は、破裂音が促音に後続する音連続を含む有意味語 10 語で（表 2)、調査方式は単語単独の読上げ式による面接録音調査である。録音はノートパソコンに直接，あるいは外付けサウンドデバイス（Edirol UA-3FX）を中継して，マイク（SONY PCM-MS957）を接続し，ノートパソコンに音声を取り込んだ（一部ポータブルリニア PCM レコーダー（SONY PCM-D1）を使用)。録音時のサンプリング周波数は 22050Hz，量子化ビット数は 16bit である。

表 2. 分析項目

有声促音項目	無声促音項目
グッバイ /guQbai/	河童（かっぱ） /kaQpa/
仏陀（ぶっだ） /buQda/	打（ぶ）った /buQta/
ベッド /beQdo/	ペット /peQto/
スラッガー /suraQgaR/	サッカー /saQkaR/
バッグ /baQgu/	バック /baQku/

音響分析に用いたソフトは、Praat (ver. 6.0.19 他)である。分析においては、先行母音、閉鎖区間、後続子音の持続時間を測定した。また閉鎖区間中に観察される voicing（を表すエネルギー）に関して、先行母音から持続し閉鎖区間中で途切れるものに関しては先行母音の終了時点からその終了時点まで、閉鎖区間の途中から始まるものに関してはその開始時点から後続子音の破裂時点までの持続時間を測定した（閉鎖区間中ずっと声帯振動が持続するものについてはこれらを区別しない)。各イベントポイント（開始点、終了点など）の認定においては、主にスペクトログラム、音声波形を用い、補助的に聴覚印象を用いた。分析のより詳しい内容については高田（印刷中）を参照されたい。

3. 結果と考察

3.1. Voicing に関する音声パターンの出現頻度

高田三枝子（2013）では促音の閉鎖区間（促音部分＋後続閉鎖音の閉鎖区間）の voicing に関する音声パターンについて次に示す 5 分類を提案し（表現は一部異なるが、内容は変わらない)、高田 (2014、印刷中) でこれを利用して地域差を観察して地域差の存在を指摘した。この分類は当初有声促音の観察に際して提案したものであるが有声性にはかかわらず促音に適用できる。本研究でもこの音声パターン分類を用い、また以下ではパターンの内容をここに示す省略形で表す。なお下記パターン中、PV はこれまで実際に観察されていない。

＜閉鎖区間中の声帯振動に関する音声パターン＞

- **声帯振動なし**(no voicing: NV)
- **声帯振動あり**
  - a. **部分的**
    - **ア**. **先行母音からの持続のみ**(remnant: R)
    - **イ**. **後続有声音の** prevoicing **のみ**(prevoicing: PV)
    - **ウ**. **(2)と(3)の不連続な併存**(R&PV)
  - b. **閉鎖区間中完全持続**(full voicing: FV)

本研究の熊本方言話者の有声促音、無声促音の発話について上記のパターンに分類した。なお先行母音からの持続的な voicing については 10msec 未満のものは「なし」とした。

世代によるグルーピングでは 1 グループの人数確保のため 20 年を 1 グループとした。ただしそれでも各世代の人数には偏りがあり、特に 70－80 代の人数は男性 1 人女性 3 人と少なく、分析の解釈においては参考程度にとどめる必要がある。各世代における音声パターンの出現割合を示したのが図 1 である。図 1 では性別についても分けて集計しているが、これは結果として性別による差が見られ、これを分けた方が世代差もより明確に示されたことによる。

まず図 1 の性別に注目すると特に FV の割合がどの世代でも男性の方が女性より高い。この性別差には男性と女性の声門動作の違いに起因する可能性を指摘できる。一般に女性の声の特徴として男性の声よりも息漏れが多くまた声帯振動時の開放時間率が大きいことが知られるが（ケント, R. D. & C. リード 1996)、これは voicing に必要な声門下圧と上圧の差の減衰が早いこと、そして voicing が早く終結することにつながると考えられる。つまり女性の方が voicing の長い持続の必要な FV を生じにくいという可能性を指摘できる。この推測が正しいとすれば、この男女差の傾向は地域に関わらず見られるはずで、今後検証する必要がある。

一方、世代差については、各性別とも上の世代ほど FV の割合が多い傾向が見られる。これは加齢による変化と世代的特徴のどちらを表すものであるかという問題がある。加齢による生理的変化としては声帯の萎縮がまず考えられ、これに関しては、声帯委縮に伴い発声時声門の閉鎖不全が生じやすくなり、声門下圧が十分に上がらない場合があることが指摘されている（角田晃一 2016)。これに従えば上の世代ほど voicing は長く続きにくくなることが予想される。しかし図 1 でみられる世代差の内容は、むしろ上の世代ほど voicing が長く続く音声が多い傾向を見せており、この予想に反する。このことから、熊本方言話者に見られる世代差は加齢によるものとは考え難く、世代的特徴を示すものではないかと思われる。

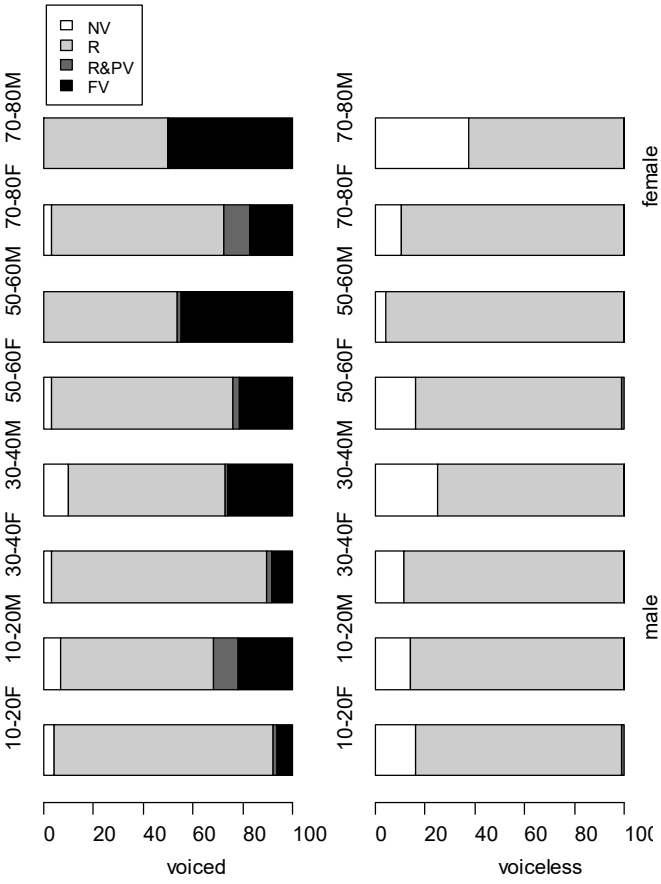


図 1 促音閉鎖部分の音声パターン比率(%)（性別×世代×有声性）

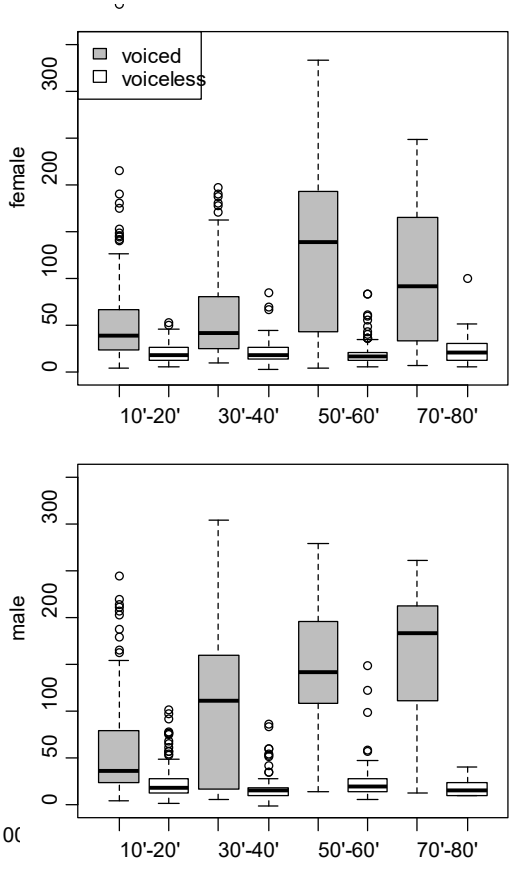


図 2 先行母音からの持続的 voicing の持続時間(msec)（有声性×世代×男女）

3.2. 先行母音からの持続的 voicing

先行母音からの持続的な voicing（以下、持続的声帯振動）の持続時間に関して、高田（2014、印刷中）では九州が他地域に比べ、長い傾向があることを指摘した。ここでは九州（熊本）内での世代差、また先の音声パターンに関して性別差が見られたため性別にも注目する。図 2 は持続的声帯振動の持続時間を、有声性別、世代別、そして男女別に示したものである [1]。

[1] 閉鎖区間中完全に voicing が持続する音声（FV）に関しては、仮に持続時間を閉鎖区間と同じとしたうえでデータとして取り入れている。持続的声帯振動は長く続けば最終的に後続音につながる。一方、後続音の prevoicing としての voicing もその開始点が最も遡れば先行母音につながる可能性もある。つ

図 2 を見ると、無声促音においては目立った世代差は見られないが、有声促音においては、一部を除いて、上の世代ほど持続的声帯振動が長くなるという傾向が見いだせる。女性 70 代以上だけがこの傾向から外れるが、これは話者人数が影響した可能性もある。性別差に関しては、全体に男性の方が持続的声帯振動が長く続く発話が多いようである。この傾向については 3 要因（有声性×世代×性別）の分散分析で、有声性と世代については 0.1%水準、性別については 5%水準で有意差が認められた。

3.3. 閉鎖区間長比率

話中（含、語中）環境の破裂音の有声性の弁別に関しては、杉藤美代子・神田靖子（1987）において、先行母音に対する閉鎖区間の比率が重要であるという指摘がある。高田（2014、印刷中）では、閉鎖区間長の「先行母音＋閉鎖区間長」に対する比率（以下、閉鎖区間長比率）について各地域とも有声音＜無声音という関係があることと、地域差があることを指摘した。特に九州は有声性による差が他地域に比べ大きかった。ここでは熊本話者内の世代と性別による分布を見ることにする。

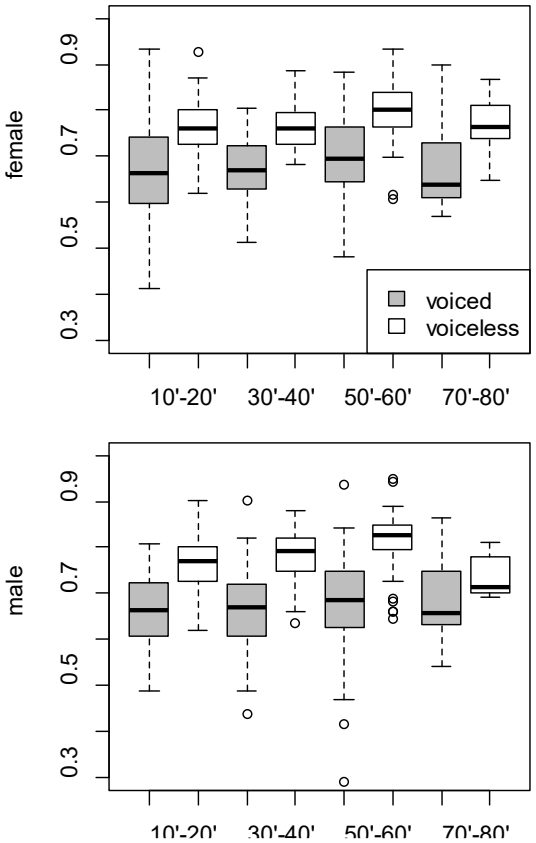


図 3 閉鎖区間長比率
（有声性×世代×男女）

図 3 は上記、閉鎖区間長比率を各世代、性別ごとに有声性に関して比較したものである。全体に閉鎖区間長比率は有声音＜無声音の関係にある。四分位範囲を見ると、70－80 代以外では、男女とも上の世代ほど有声音と無声音の分布域(四分位範囲で見て)が上昇している。3 要因の分散分析においても、有声性と世代については 0.1%水準で有意差が確認される。一方、性別については確認されない（p=.89）。交互作用は性別と有声性の間でのみ 0.5%水準で確認された。分布範囲が上の世代で上昇することについては発話速度あるいは単語読み上げという調査スタイルが影響した可能性も考えられる。ここではむしろ有声／無声間の分布の重なりに注目したい。これに関して、70－80 代男性以外は、各グループとも重なりは小さく有声性により閉鎖区間長比率が明確に異なる。ただし高田（2014、印刷中）で報告した九州の結果ほど有声性による分布の開きはない。これに近いのは今回の結果では 30－40 代、また 50－60 代男性である。なお有声性による分布

まり FV はその voicing 区間のどの程度をこの先行母音からの持続的な voicing、あるいは後続音の prevoicing が占めているか厳密にはわからない（エネルギーの減衰と増大から推測できる可能性はある）。ただし九州に関してはあまり後続子音の prevoicing が頻繁にそして長くは観察されないことと、またこの FV の音声を分析から除くことで特に高年層の持続的声帯振動の分析が難しくなるためこのような措置とした。

範囲の境界はいずれのグループのおおよそ 0.7～0.8 あたりにあり、これは高田（2014、印刷中）の報告とほぼ一致する。

70－80 代で閉鎖区間長比率の分布範囲が他世代に比べ少々小さくなっていることについては話者数の少なさの影響も考えられるが、あるいは一つの可能性としてシラビーム方言的な発音の影響も考えられるかもしれない。秋山（1983）によれば、熊本方言は基本的にはシラビーム方言ではないが、部分的に特殊拍が短めに発音される現象が、特に南部を中心に観察されるという。今後、発話速度との関係を含め検討したい。

4. 本研究の成果と今後の課題

以上、熊本方言において、促音の音声詳細に関しては世代差が存在することが明らかになった。そして本研究でみられた世代差は、加齢より世代的特徴として考えるべきではないかと提案した。そうであるとすれば、熊本方言の古い世代の姿、すなわち伝統的方言としては、有声促音では FV となるほど閉鎖区間中声帯振動が頻繁に長く続けられ、無声促音では振動しないという明確な違いのある発音であったものが、現在、有声促音でも閉鎖区間中あまり声帯振動がないという発音に変わりつつあるとみることができる。また性別による明確な違いも報告した。これも今後他の地域の観察の際にも考慮する必要がある。なお特に閉鎖区間長比率に関しては発話速度(あるいは単語単独発話というスタイル)による影響も考えられ、他スタイルの資料と比較するなどの検証が必要である。なお近く大阪、東京、秋田といった他地域についても本稿と同様の観点からの報告を予定している。

【謝辞】本研究は財団法人博報児童教育振興会「2005 年度第 1 回博報『ことばと文化・教育』研究助成」による資料を用い、また JSPS 科研費 15K16762 の助成を受けて進めています。本稿執筆の際、松井理直氏から有益なコメントをいただきました。記して感謝いたします。

【引用文献】

秋山正次(1983)「熊本県の方言」飯豊毅一・他（編）『講座方言学９　九州地方の方言』, 207-235p. 国書刊行会.

角田晃一(2016)「高齢者の嗄声への対応」『日本耳鼻咽喉科学会会報』119-11, 1442-1443p.

ケント, R. D. & C. リード（荒井隆行・菅原勉(監訳)）(1996)『音声の音響分析』, 海文堂.

松浦年男（2016）「天草諸方言における有声促音の音韻論的・音声学的記述」『国立国語研究所論集』10, 159-177p.

杉藤美代子・神田靖子(1987)「日本語話者と中国語話者の発話による日本語の無声及び有声破裂子音の音響的特徴」『大阪樟蔭女子大学論集』24, 1-17p.

高田三枝子(2011)『日本語の語頭閉鎖音の研究―VOT の共時的分布と通時的変化』くろしお出版.

高田三枝子(2013)「有声破裂音の後続する促音閉鎖区間の有声性に関する音声パターン」『明海日本語』18(増刊), 15-30p.

高田三枝子(2014)「有声促音の音声的有声性に見られる地域差」『日本音声学会第 26 回全国大会予稿集』, 165-168p.(発表スライドの公開 URL(2017 年 7 月現在)http://researchmap.jp/yearman/voicedgeminateWS/)

高田三枝子(印刷中)「促音閉鎖区間の有声性に関する音声詳細の地域差」『人間文化』32, 愛知学院大学人間文化研究所.

高山倫明(2012)「第 5 章　促音の音用論」『日本語音韻史の研究』, 129-145p. ひつじ書房.

C3

近畿方言話者における母音の無声化—言語意識と言語選択—

邊姫京（国際教養大学）
byun@aiu.ac.jp

1. はじめに

近畿方言における無声化生起率は，東京方言のそれに比べるとかなり低いが（杉藤 1988，藤本・桐谷 2003，安田・林 2011，邊 2010，2012 など），近畿方言の中でも世代の差があり，若年層の無声化生起率は中年層，高年層のそれより高い（邊 2010，2011，2012）．

近畿方言の若年層における無声化生起率の増加は，共通語化の影響と見られるが，共通語の普及により地域によっては方言が話せない若者も増えている．方言が話されている地域では相手や場面により方言と共通語を使い分けているとされるが，若者の場合，調査時に話されたことばが使い分けられた共通語なのか，完全に置き替えられた共通語なのか判断が難しいときがある．

安田・林（2011）は，20-30 歳代の近畿方言話者 10 名に，単語（大学生，美しい，包む，季節，深い，寿司，飛行機，子孫，少し，二人，敷く，の 11 語）と，文章（上記の 11 語を含む文章）を読ませ，無声化生起率を調べている．平均無声化生起率は，単語と文章の順に，東京方言話者は 80%，92%，近畿方言話者は 24%，73%で，単独発話において東京方言と近畿方言の差が大きい．単独発話のほうが文章よりも丁寧に読まれる傾向はあり，それが無声化生起率に影響した可能性もあるが，主な要因は音環境と音調にあると思われる．音環境は，近畿方言に限らず全国的に摩擦音に挟まれた狭母音はそれ以外の環境に比べて無声化が相対的に起こりにくいが（邊 2012），上記の 11 語のうち 2 語がそうである．音調は，上記の 11 語を京阪式アクセントで読むと無声化拍の音調が H，後続拍の音調が L になる語が 5 語もある．近畿方言の若年層において無声化拍と後続拍の音調が HL の場合は LL，LH，HH の場合より無声化が有意に起こりにくく（邊 2011），音調の影響で無声化生起率がさらに低くなったと思われる．音調が HL であったとすれば，話者は共通語ではなく方言で読んだことになるが，論文には収録時の指示内容についての言及がないので，話者が共通語と方言のどちらで読んだかは不明である．

そこで本稿では，具体的な指示を与えて単語と文章を読んでもらい，方言と共通語とで無声化生起率にどの程度の差があるかを調べた．方言で読む単語の単独発話は，前後の子音，母音の種類など音環境をそろえた．また，京阪式アクセントで無声化拍と後続拍の音調が HL にならない語を選び，アクセントによる無声化生起率への影響を最小限にした．加えて，話者の言語意識を知るために，特定の場面で関西弁と共通語のどちらを使うかを問うアンケートを実施した．

さて，近畿方言における無声化生起率の世代差は，その結果が今後も持続する通時的変化（diachronic change）と見られるが，共時データに見られる世代差を通時的変化と見るに

は，二時点以上の調査結果を比較し，同一コーホート（cohort）において無声化生起率に変化がないことを確認する必要がある．一時点の調査で観察される世代差から通時的変化を推定する見かけの時間（apparent time）の推定法は，通時的変化を成長に伴いことばが変化するエイジ・グレイディング（age-grading）と見誤る恐れがあり，エイジ・グレイディングの可能性を排除する必要があるからである．二時点以上の調査結果を比較し，値に変化がなければ通時的変化，変化があればエイジ・グレイディングと解釈できる．

本稿の目的は二つある．一つは，近畿方言の若者において無声化生起率が東京方言に比べて低いことが，無声化を習得していないことに起因するのか，あるいは習得しているが意図して無声化せずに発話していることに起因するのかを，言語意識と関連して探ることである．もう一つは，1990 年代生まれの話者の無声化生起率が，10 年前に同地域，同方言話者を対象に行った結果（邊 2012）と比較して差があるかを調べることである．

2. 手順

2.1. 話者と音声収録

話者は, 大阪または兵庫で生まれ育った男女 23 名である（男性 3 名, 女性 20 名, 1990-1997 年生まれ）．音声収録は，2017 年 1 月に神戸大学内で行われた．

2.2. 音声資料

音声資料は，単独発話（単語リスト）と文章（朗読文）の 2 種類である．表 1 に単語発話 36 語の音環境を示す．下線の拍が無声化拍である．各セルは，無声化拍の母音が上段 : /i/，下段 : /u/，後続拍の母音が左側 : 非狭母音/a/または/e/，右側 : 狭母音/i/または/u/である．後続子音が破擦音の場合は，後続母音が狭母音のみで非狭母音はない．音声収録の際は，無声化が起こらない環境の語をダミーとして入れ，計 76 語をランダムに並べた単語リストを使用した．単語リストの上段には「1 つずつ区切って，普通の速さで，地元で家族や親しい友人に話すように自然に読んで下さい」という指示文があり，同様の内容を口頭でも説明した．

語アクセントは主に『京阪系アクセント辞典』（中井幸比古編 2002，勉誠出版）を参照したが，『全国アクセント辞典』（平山輝男編 1960，東京堂）も参考にした．無声化拍と後続拍の音調が HL になる語を避け，LH，HH，LL になる語を選んだが，話者によっては HL で発話する例もある．

表 1: 単語発話の音環境

後続 / 先行	破裂音		破擦音		摩擦音	
破裂音	きたかみ（北上）	ききめ（効き目）	あきち（空き地）	きつね（狐）	きせい（既成）	きすぎる（着過ぎる）
	くけい（矩形）	くくる（括る）	くちびる（唇）	たいくつ（退屈）	くさり（鎖）	くすり（薬）
破擦音	ちかく（近く）	ちくわ（竹輪）	ちち（父）	ちつじょ（秩序）	ごちそう（御馳走）	ちしりょう（致死量）
	おつかい（お使い）	つきあい	つち（槌）	いつつ（五つ）	あつさあたり	うつし（写し）
摩擦音	しかい（市会）	いしき（意識）	しちご（七五）	しつもん（質問）	しさつ（視察）	ししまい（獅子舞）
	ふたご（双子）	いふく（衣服）	スチーム	ふつう（普通）	ふせい（不正）	ふしぎ（不思議）

朗読用の文章は「これは，私が小さいときに，村の茂平というおじいさんからきいたお話です」から始まる『ごん狐』（新美南吉作）の冒頭 16 文である．表 2 に分析対象になる語の音環境を示す．該当する語がないセルは空白になっている．下線の拍が無声化拍である．無声化拍が連続する「すすき」は 2 回登場する．表 2 は語中の 29 語で，他に文末の「です」「ます」が 4 回，「でした」「ました」が 12 回登場し，分析したのは計 45 語である．音声収録の際は「学校で朗読するつもりで読んでください」と口頭で指示を出した．語アクセントは話者によって東京式になる場合も京阪式になる場合もあるが，そのまま分析した．

表 2: 朗読文の音環境

後続 / 先行	破裂音	破擦音	摩擦音
破裂音	出られなくて, きて	南吉（なんきち）, きつね	ひゃくしょう（百姓）, くさ
破擦音	ちかく, つけたり, つかる, 見つからない	つつみ（堤）	
摩擦音	わたしたち, すこし, ひとり, そして, ちらしたり, ほして, つるして, ほっとして, ひかって, すくない, 増して, ふと, ひと, ふかい, すすき1, すすき2		すすき1, すすき2

2.3. 言語意識に関するアンケート

音声収録の前に，A から F について「次の状況であなたが使用するのは関西弁と東京弁のどちらですか？」を尋ねるアンケートを実施した．アンケート用紙には「どちらかに○をしてください．両方の場合は両方に○をしてください.」という指示文がある．

A 家族と話すとき　　B 学校で（関西出身の）友人と話すとき
C 学校で（関東出身の）友人と話すとき　　D みんなの前でプレゼンをするとき
E 知らない人に東京弁で声をかけられたとき　F NHK に出演してインタビューを受けるとき

2.4. 分析

収録した音声は音響分析ソフトを用いて無声化の有無を判断した．音声波形に周期的な波形が現れず，広帯域スペクトログラム上にも声帯音源パルスが現れない場合を無声化母音とし，その母音の割合を求め無声化生起率とした（邊 2011）．無声化生起率は個人ごとに算出した．単独発話は，無声化拍と後続拍の音調が HL かどうかを，筆者が聴覚判断した．

3. 結果

3.1. 無声化生起率

結果を単独発話・文章，音環境の順に提示する．話者のほとんどが女性なので男女は分けない. 大阪と兵庫とで有意差は見られないので（$t(21)=0.2554, p=0.8009$）地域も分けない.

図 1 に単独発話（方言バージョン）と文章（共通語バージョン）の結果を箱ひげ図で示す．箱内の太線は中央値，図下の有効数は話者数である．無声化生起率の平均は，左から「単独・語中」41%，「文章・語中」78%，「文章・文末」100%である．文末は 16 語すべてが無声化しているので図 1 では中央値の 100%のみ表示されている．単独発話と文章の語中に対して，対応のある t 検定を行った結果は，文章のほうが単独発話より有意に無声化生

起率が高い（t(22)=11, p<0.0001）.

本稿と安田・林（2011）の結果を平均で比較すると，「文章・語中」は，安田・林が 73%，本稿が 78%で非常に近い．「単独・語中」は安田・林が 24%，本稿が 41%で本稿のほうが高い．前述したように，安田・林（2011）の調査語には東京式アクセントで発話された場合は無声化が起こりやすいが，京阪式アクセントで発話された場合は無声化が起こりにくい語が多数含まれている．そのため本稿よりも無声化生起率が低くなったと思われる．

本稿の単独発話 828 例（23 名×36 語）のうち，無声化拍と後続拍の音調が HL で発話された語は 79 例で，うち 3 例が無声化し，音調が HL の場合の無声化生起率は 4%である．これに対して無声化拍と後続拍の音調が HH，LL，LH の場合の無声化生起率は 45%である（749 例のうち 334 例が無声化）．イエーツの補正を用いたカイ二乗検定を行った結果は，音調が HL の場合はそれ以外の場合より無声化生起率が有意に低い（$x^2(1) = 47.6, p < 0.0001$）.

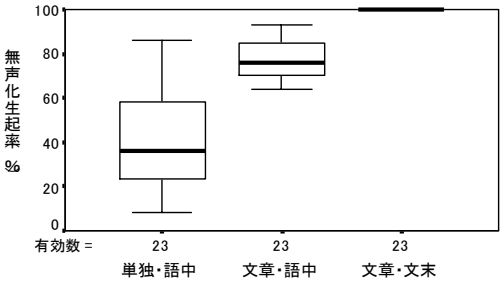


図 1: 単独発話（方言）と文章（共通語）の無声化生起率

音環境のうち子音別，母音別結果は紙幅の都合により省略し，前後子音別結果のみ提示する．図 2 に前後子音別結果を示す．右側の「文章」は，破擦音－摩擦音，摩擦音－破擦音に該当する語がないためグラフは表示されていない．単独発話と文章のいずれも前後子音が摩擦音の場合は無声化生起率が極めて低い．単独発話で破擦音－閉鎖音，摩擦音－破擦音，摩擦音－閉鎖音の無声化生起率が相対的に高いのは，邊（2012）と同じ結果である．

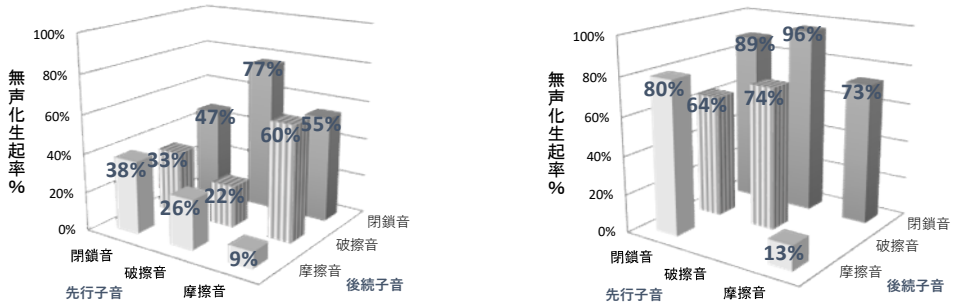


図 2: 単独発話（左）と文章（右）の前後子音別無声化生起率

3.2. 言語意識

図 3 に特定の場面で関西弁または東京弁（共通語）のどちら（または両方）を使うかを尋ねたアンケートの結果を示す．概して家族や友人と話すウチまたは私的な場面（A/B/C）

では関西弁，ソトまたは公的な場面（D/E/F）では共通語という使い分け意識が働いていると言える．しかし，E や F のように共通語が期待される場面でも関西弁を選択した人が半数を超えており，公的な場面における共通語への切り換えは十分とは言いがたい．D のみんなの前でプレゼンをするアカデミックな場面では，共通語の選択が半数を超えている．

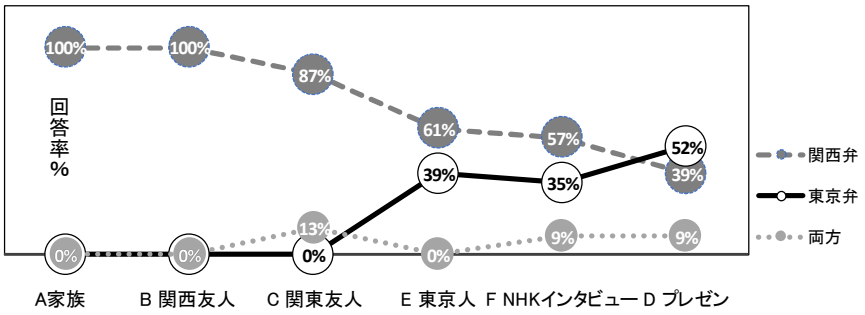


図 3: 言語意識に関するアンケート

3.3. コーホートの無声化生起率

図 4 に本稿の結果（2016 年度調査）と，10 年前に同じく 1990 年代生まれの大阪・兵庫出身の話者 23 名を対象に行った調査の結果（2006 年度調査）を示す．2006 年度調査の発話資料（邊 2010，2012）は，共通語アクセントを基準にしている．また，発話の際の指示文には「1 つずつ区切って，ふつうの速さで，いつも言うように，自然に読んでください.」とあるが，どちらかと言えば，共通語で読まれた可能性があり，本稿とは調査方法で多少違いがある．平均無声化生起率は，2006 年度調査が 42%，2016 年度調査が 41%で，有意差は認められない（t(44)=1.23, p=0.2254）．同一コーホート（1990 年代生まれ）の無声化生起率は 10 年の歳月が経ても有意な差は見られず, 無声化生起率は維持されていると言える．

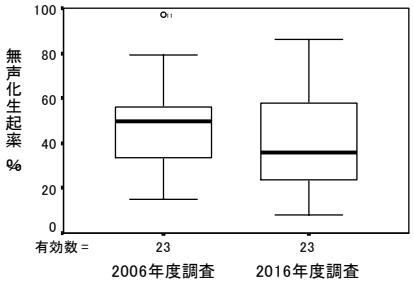


図 4: コーホートの無声化生起率

4. まとめと考察

1990 年代に大阪・兵庫に生まれ，その後も同地で過ごした近畿方言の母語話者 23 名を対象に，母音の無声化について，方言と共通語とでどの程度の差があるか検討した．加えて，言語意識の知るために特定の場面で関西弁と東京弁（共通語）のどちらを使うか尋ねるアンケートを実施した. さらに, 本稿の結果と 10 年前にほぼ同様の方法で得た結果を比較し，同一コーホート（1990 年代生まれ）の無声化生起率に差があるかを調べた．

平均無声化生起率は，単独発話（方言）が 41%，文章（共通語）が 86%で，共通語で発話した場合は方言の場合に比べて無声化生起率が有意に高い. 朗読用の文章には文末に「デス，マス，シタ」が 16 回出現しているが，これら 3 語の無声化生起率は 100%，これら 3 語を除いた場合でも 78%である．邊（2010）によれば，従来無声化が目立つとされる地域の無声化生起率は平均 60%以上，東京方言のように無声化が規則的に起こるとされる地域の無声化生起率が平均 80%以上で，今回の話者グループは，無声化生起率に関して，共通語能力をほぼ獲得していると言える.

言語意識に関しては，家族や友人と話す私的な場面ではほとんどが関西弁を好むが，知らない人から東京弁で声をかけられる，NHK に出演しインタビューを受けるといった公的な場面においても半数以上の人は関西弁を好んでいる．ただし，みんなの前でプレゼンをする場合は半数以上が東京弁（共通語）を選択しており，アカデミックな場面では書き言葉のように共通語意識が強く働くようである.

上記の無声化生起率と言語意識に関する結果を総合すると，今回の話者グループは，無声化に関して，共通語は話せるが，意図して話さないという状況が浮かび上がってくる.

さて, 1990 年代コーホートの無声化生起率は, 10 年の歳月を経ても有意な差は見られず, 近畿方言の若者に見られる無声化生起率の増加は通時的変化と見られる．この結果を言語変化の S 字カープモデルに当てはめれば，共通語の場面では今後も無声化生起率が増加することが予想されるが，その一方で，意図的に共通語を使用しない意識が強まれば，無声化生起率の増加は見込めない．近畿方言の若者の無声化生起率が音調に左右されるのは明らかに共通語の影響であり，従来の近畿方言の無声化には見られない特徴である．共通語化，言語意識と関連して，近畿方言の無声化については，今後もその動向が注目される.

謝辞

音声収録の際にご協力くださった話者の皆さま，録音場所や話者のリクルート等にご尽力くださった神戸大学の林良子先生に深く感謝申し上げます.

参考文献

杉藤美代子（1988）「日本の 8 都市における母音の無声化」『大阪樟蔭女子大学論集』25, 1-10.

邊姫京（2010）「5 地域に見る無声化生起率の年齢的変化」『日本語の研究』6(4)，79-94.

邊姫京（2011）「京阪神における狭母音の無声化」『音声研究』15(2)，23-37.

邊姫京（2012）「日本語狭母音の無声化―共通語普及の指標として―」博士学位論文，東京大学.

藤本雅子・桐谷滋（2003）「東京方言と近畿方言における無声化の比較」『音声研究』7(1), 58-69.

安田麗・林良子（2011）「日本語学習者における母音無声化―台湾人日本語学習者，東京，近畿方言話者を対象に―」『音声研究』15(2)，1-10.

C4

ムラブリ語の音節とソノリティ階層

伊藤雄馬 (日本学術振興会・東京外国語大学)
yuma.i.86@gmail.com

1　はじめに

ムラブリ語（オーストロアジア語族クム語派、タイ・ラオス）の音節は、主音節（major syllable）と副音節（minor syllable）に分けて記述されてきた（Rischel 1995、伊藤 2014、Bätscher 2015）。この記述方法は東南アジア地域の言語記述、特にオーストロアジア語族の記述によく見られる（cf. Jenny & Sidwell 2015）。ムラブリ語の場合、例 1 の下線部分が副音節となる。

(1)　**hn.taʔ** “tail” （Rischel 1995: 261）

kl.duɯh “uvula” （Rischel 1995: 276）

tr.nɛk “hummer” （Rishcel 1995: 324）

主音節は母音が音節核を担い、単独で現れうる。一方、副音節は子音が音節核を担い、必ず主音節に先行する。それぞれの音節の特徴を表 1 にまとめる。

表 1: 主音節と副音節

	音節核	現れる環境
主音節	母音	自由
副音節	子音	主音節の前

伊藤（2014: 43）では、主音節と副音節の特徴を踏まえ、ムラブリ語の音節テンプレートを図 1 のように提示し、主音節と副音節の各スロットに入る子音を表 2、3 のようにまとめている。なお表にある阻害音とは、閉鎖音と摩擦音の両方を含む自然類である。音素目録は付録を参照されたい。

$$\overbrace{C_{-2}C_{-1}}^{\text{(副音節)}}\overbrace{C_1(C_2)(C_3)V(C)}^{\text{主音節}}$$

図 1: 伊藤（2014: 43）の音節構造テンプレート（一部改変）

表 2: 主音節の頭子音連続の例（伊藤 2014: 47 より一部改変）

C_1	C_2	C_3	例	
阻害音	震え音	-	preʔ	「トウガラシ」
阻害音	鼻音	-	hmɛʔ	「新しい」
阻害音	側面音	-	pleʔ	「実」
阻害音	-	接近音	pyɛ	「トカゲの一種」
鼻音	震え音	-	mrek	「絞る」
鼻音	側面音	-	mlaʔ	「人」
鼻音	-	接近音	myɤ	「妻」
阻害音	震え音	接近音	grwɛc	「爪」
阻害音	側面音	接近音	klwen	「方向」

表 3: 副音節（伊藤 2014: 47 より一部改変）

C_{-2}	C_{-1}	例	
阻害音	鼻音	sm.bɛp	「口」
震え音	鼻音	rm.bah	「側面」
阻害音	震え音	kr.poʔ	「稲光」
阻害音	側面音	kl.muy	「毛」

ここまでの図や表を見て分かるように、ムラブリ語の音節にまつわる記述は煩雑であり、より簡潔な記述方法が求められてきた。

本稿では、「副音節」を子音連続として再解釈し、また接中辞（infix）の振る舞いを観察することによって、音節をより簡潔に記述する方法を提案する。また、この議論から明らかとなるムラブリ語のソノリティ階層の特徴についても論じる。

2 「副音節」の再解釈

ムラブリ語の語彙データベースである伊藤（2017）から、「副音節」を含む形式を 63 個抽出した。これら「副音節」を子音連続として再解釈し、分類したのが以下の表 4 である。

表 4: 「副音節」を含む形式

子音連続数	数	例
CC-	0	
CCC-	61	**smbɛp**「口」、**rmbah**「側面」、**bnliŋ**「青系の色」、**prpip**「(間口が) 広い」、**sŋkɛr**「爪」、**srbok**「蹄」、**srkɛŋ**「羽; 鶏」、**srleŋ**「壁」、**srmal**「魂」、**srɓot**「歌う」、**crluh**「差し込む」、**drmɔʔ**「イタチの一種」、**drtʰaŋ**「竹の一種」、**gmtak**「舌打ちする」、**kndɯl**「尻」、**gnrɛ**「スープ」、**krlaj**「足首、手首」、**krwɛc**「指」、**hnda?**「薄い」、**hnta?**「尾」、**hnrɤʔ**「成人女性」、**hŋkok**「銃」、**hŋkeʔ**「薪: ダニ」、**hŋgik**「かぼちゃ」、、**hrlah**「捌く」、**hrliŋ**「忘れる」、**hrlɛʔ**「笑う」、**hrloj**「流れる」、**ɲrak**「櫛」、**ɲrmɯk**「嫌う」、**kldul**「太い」、**kldɯl**「切り株」、**klkil**「膝」、**klcol**「肘」、**klmuj**「毛」、**kndiŋ**「臍」、**kndɔʔ**「屋根」、**krbol**「埃」、**krlap**「又木」、**krpoʔ**「稲光」、**krpel**「とんぼ」、**krwac**「箒」、**krwɛc**「引っ掻く」、**krwen**「方向」、**lmbaʔ**「牛」、**lmŋɔr**「蜂の一種」、**pmpo**「象」、**prgʌr**「磨く」、**prlɯh**「唸る」、**rŋgap**「顎」、**tʰldil**「夕闇」、**tʰlkal**「白髪」、**tʰrbɛʔ**「羊」、**trlak**「傷」、**trlam pam**「苔」、**trlaŋ**「鷹」、**trlɔh**「鍋」、**trluh**「竹の一種」、**trnap**「大きい」、**trnɛk**「鎚」、**trnɯj**「臼」、
CCCC-	2	**krkrwɛc**「引っ掻く」、**bmbwaj**「蜘蛛」
CCCCC-	0	
計	63	

「副音節」を含むとされる形式の約 97%（61/63）が 3 子音連続であることが分かった。4 子音連続は 2 例のみで、どちらも重複が関係している（**krkrwɛc**「引っ掻く」、**bmbwaj**「蜘蛛」：下線部が重複部分）。加えて、「副音節」とされる形式には、2 子音連続と 5 子音連続は存在しないことも明らかになった。5 子音連続については、伊藤（2017）の全ての語彙の中においても観察されなかった。

つまり、「副音節」を含む形式は（重複を含む例を除けば）全て 3 子音連続であることが分かる。

3　接中辞の振る舞い

ムラブリ語には接中辞が存在する。そのひとつに**-rn-**がある。この接中辞は動詞について、その動詞に関連する名詞を派生させる。この接中辞は、いくつか異形態を持つ。その分布は、語根の頭子音（onset）によって予測可能である（伊藤 2014: 57、Bätscher 2015: 1009）。例（2）を見られたい。頭子音単独の場合、**-rn-**が頭子音の後に挿入される。震え音を含む子音連続の場合、**-n-**が子音連続の間に挿入される。それ以外の子音連続の場合、**-r-**が子音連続の間に挿入される。なお、3 子音連続に接中辞が挿入される例は、観察されていない。

(2)　**t-rn-ɛk**「鎚」 vs. **tɛk**「叩く」
ɟ-n-rak「櫛」 vs. **ɟrak**「髪を梳く」
k-r-wac「箒」 vs. **kwac**「掃く」

注目したいのは、この接中辞によって派生した形式は、語根の頭子音の数に関わらず、必ず 3 子音連続になる点である。この観察は「副音節」を含む形式が全て 3 子音連続であったことと符合する。

4　音節構造の再モデル化

以上までの議論でわかったことをまとめる。

(3)　a.「副音節」を含む形式の約 97%が 3 子音連続
b. 4 子音連続は、重複が関与した 2 例のみ
c. 5 子音連続は観察されない
d. 接中辞**-rn-**により派生した形式は、語根の頭子音数に関わらず必ず 3 子音連続となる

このうち a. と d.、つまり「副音節」を含む形式と接中辞**-rn-**によって派生した形式の全てが 3 子音連続となる事実は、ムラブリ語において最大頭子音連続は 3 であることを示唆しているものと考える。

(4)　a.+ d. 最大頭子音連続は 3 である

以上から、図 1 の音節テンプレートをもう一度評価してみる。まず、図 1 のテンプレートでは、最大子音連続が 5 子音連続に見える。しかし、実際は 5 子音連続は観察されず、最大子音連続は 3 と考えられるため、不適切である。また、そもそも副音節と主音節という二つの種類の音節を想定する根拠がなく、アドホックである。

この問題を解決するために、本稿では新たに図 2 のような音節テンプレートを提案する。

C(C)(C)V(C)

図 2: 本稿の提案する音節テンプレート

図 2 の音節テンプレートでは、アドホックであった副音節と主音節の区別を廃した。また、3 子音連続が最大であることも明確である。なお、重複が関与する場合のみ 4 子音連続を許す理由は別稿に譲ることとする。

5　ソノリティ階層

ところで、ムラブリ語の 2 子音連続（cf. 表 2）から、図 3 のようなソノリティ階層を提示できる。図 3 では、閉鎖音と摩擦音を阻害音、震え音と側面音を流音としてまとめてある。

阻害音＜鼻音＜流音＜接近音＜母音

図 3: ソノリティ階層

新しい音節テンプレートをもとに、ムラブリ語の語彙を見ていくと、これまで副音節とされてきた形式は、ソノリティ階層に準ずる形式と違反する形式に分けることができる。ソノリティ階層に違反する形式は、表 5 に示す 4 通りに限られることが分かった。

表 5: ソノリティ階層に違反する形式

阻害音	鼻音	阻害音	**smbɛp**	「口」
流音	鼻音	阻害音	**rmbah**	「側面」
阻害音	流音	阻害音	**krpoʔ**	「稲光」
阻害音	流音	鼻音	**klmuj**	「毛」

表 5 は 1 節の表 3 と概ね対応するが、表 5 は「副音節」の後に続く子音まで含む点で表 3 よりも優れた一般化だと言える。

従来の分析では、（明示的に述べられていはいなかったものの）ソノリティ配列原理（Sonority Sequencing Principle）に従って音節を定めていた。そのため、ソノリティ階層に反する部分は「副音節」として音節に認定する必要があった。

本稿の分析では、ムラブリ語の音節がソノリティ階層に違反しうると考える。その結果、「副音節」を認める必要がなくなり、音節構造のテンプレートを単純化できた。加えて、子音連続をソノリティ階層に準ずる音節と違反する音節に分けることで、より簡潔で説明力の高い一般化が行えることが分かった。

5.1　声門閉鎖音を含む子音連続

ソノリティ階層に違反する形式を認めた結果、声門閉鎖音を含む形式の分析に影響を与える。声門閉鎖音を間に含む CV_1CV_2-の形式を見てみる（例 5）。これは、V_1 と V_2 の質が常に一致する。

(5)　**baʔaʔ**「穢れている」

tʰuʔur「暑い」

lɔʔɔk「棘」

つまり、声門閉鎖音の直前の母音 V_1 は予測可能であり、音韻的ではないと解釈できる。ただし、これまでソノリティ階層に反する音節を認めていなかったため、音節構造のテンプレートに従って解釈し、予測可能な母音である V_1 も音韻的であると考えてきた。

本稿ではソノリティ階層に違反する音節をムラブリ語に認めた。つまり、これらの形式もソノリティ階層に違反する音節として記述できる。予測可能な母音は、ソノリティ階層違反の代償で現れると考えられる。よって、音韻的な解釈を以下のように改める。

(6)　**baʔaʔ → bʔaʔ**「穢れている」

tʰuʔur → tʰʔur「暑い」

lɔʔɔk → lʔɔk「棘」

例（6）のような形式を、声門閉鎖音を含む 2 子音連続と再解釈した。本節の最初に述べたように、ソノリティ階層は 2 子音連続の観察に基づいて定めてある。声門閉鎖音を含む形式を 2 子音連続と再解釈したため、この解釈はソノリティ階層に影響を与える。つまり、阻害音-声門閉鎖音の連続は、代償の母音が現れることから、ソノリティ階層に違反する子音連続である。そして、声門閉鎖音は、ソノリティ階層において、阻害音から独立して、阻害音よりも低い位置にあると考えられる。

声門閉鎖音＜阻害音＜鼻音＜流音＜接近音＜母音

図 4: 新しいソノリティ階層

ソノリティ階層について論じた Parker (2002: 225-6) によれば、英語の声門閉鎖音は音響音声学的に共鳴音と阻害音の両方の特徴を持つため、ソノリティ階層内の位置づけが困難であるとしている。それゆえ、声門閉鎖音は、言語によってソノリティ階層内の位置が異なる可能性があり、声門閉鎖音が単独の音素として認められる言語における研究が必要であるとも述べている。

ムラブリ語は声門閉鎖音を単独の音素として持つ（付録参照）。またソノリティ階層内において、声門閉鎖音が阻害音より低い位置にあることが、本節の議論から示せる。声門閉鎖音のソノリティ階層の位置付けにおいて、ムラブリ語は類型的に有意義なデータを提供するものと思われる。

6　分析の応用

本稿の分析が他の言語、特にオーストロアジア語族クム語派の言語に応用できる可能性がある。例えば、クム語（オーストロアジア語族クム語派）に、3 節で見た接中辞**-rn-**の振る舞いと同じように見える例が散見される。

(7)　クム語（Svantesson 1983: 96-7)

p-rn-aam ‘trap’ vs. **paam** ‘to set a trap’

p-r-lis ‘plectrum’ vs. **plis** ‘to pluck’

Svantteson (1983: 12) はクム語の音節構造に副音節と主音節を区別している。本稿のムラブリ語の分析を応用すれば、その区別なしに音節を記述できる可能性がある。

7　まとめ

本稿の主張を以下にまとめる。

(8)
- ムラブリ語の音節は、音節核が母音で、最大頭子音連続は 3 である
- ソノリティ階層に違反する音節を認め、副音節と主音節の区別を廃することができる
- 声門閉鎖音はムラブリ語のソノリティ階層において、阻害音より低い位置にある
- 他の言語（特にオーストロアジア語族クム語派）に本稿の分析が応用できる可能性がある

参考文献

Bätscher, Kevin (2015) “Mlabri.” In Jenny, Mathias & Paul Sidwell (Eds.), *The Handbook of Austroasiatic Languages.*, 1003–1030. Leiden, Boston: Brill.

伊藤雄馬 (2014) 「ムラブリ語の文法スケッチ」『地球研言語記述論集』6, 41–72.

伊藤雄馬 (2017) 『ムラブリ語語彙集』未刊行.

Jenny, Mathias & Paul Sidwell (Eds.) (2015) *The Handbook of Austroasiatic Languages.* Leiden, Boston: Brill.

Parker, G. Stephen (2002) *Quantifying the Sonority Hierarchy.* Doctoral dissertation, University of Massachusetts Amherst.

Rischel, Jørgen (1995) *Minor Mlabri: A Hunter-Gatherer Language of Northern Indochina*. Copenhagen: Museum Tusculanum Press.

Svantesson, Jan-Olof (1983) *Kammu Phonology and Morphology*. Lund : CWK Gleerup.

付録

ムラブリ語では、頭子音（onset）と末子音（coda）の体系が異なるため、別々に示す。頭子音には 24 の音素を認め、末子音には 14 の音素を認める。

表 6: 頭子音の目録

	両唇	歯茎	硬口蓋	軟口蓋	声門
閉鎖音	p	t	c	k	ʔ
	b	d	ɟ	ɡ	
	p^h	t^h		k^h	
	ɓ	ɗ			
摩擦音		s			h
鼻音	m	n	ɲ	ŋ	
震え音		r			
側面音		l			
接近音	w		j		

表 7: 末子音の目録

	両唇	歯茎	硬口蓋	軟口蓋	声門
閉鎖音	p	t	c	k	ʔ
摩擦音					h
鼻音	m	n	ɲ	ŋ	
震え音		r			
側面音		l			
接近音	w		j		

母音には 10 の音素を認める。母音の長短は弁別的でない。

表 8: 母音の目録

	前舌	後舌/非円唇	後舌/円唇
狭	i	ɯ	u
半狭	e	ɤ	o
半広	ɛ	ʌ	ɔ
広		a	

C5

古典語におけるモーラの位置づけの再考とモーラ・タイミング

イー・リスニング研究所
大竹孝司

1. はじめに

20 世紀の近代言語学は音声言語の探求に大きな影響を与えた２つの用語を生み出した。モーラとモーラ・タイミングである。前者は、アクセント・音節構造・語彙認識など音声言語の音韻が関与する様々な機能を説明するのに用いられてきた用語である。一方、後者は主として日本語の話し言葉の知覚上のリズムを説明するために用いられてきた用語である。この両用語のモーラの概念は古典ギリシャ語の韻文から転用されたもので、その定義は長短音節の組合せから出現する古典ギリシャ語の韻文のリズムに対して発話の観点から示した時間的な単位（timing unit）とされる。

モーラの用語が現代の音声言語の探求に取り入れられてから 4 分 3 世紀以上が経過したが、定義に関して論争が繰り広げられてきた。その多くは音韻上の解釈の問題である（詳細は Fox 2000: 46-50 参照）。一方、モーラ・タイミングは等時性を有する時間がリズムを生むという説であるが、実証研究の結果からその存在は否定された（Patel 2008: 119-122 参照)。そして、等時性に代わる新たな説が提唱された。この新説では等時的性質はヒトの発話行為の結果ではなく、当該言語の種々の音韻情報（音節構造、アクセントの分節音への影響など)の影響が最小のものがモーラ・タイミングであるとした（Dauer 1883)。この説を更に発展させ nPVI などのモデルに関心が移り今日に至る（Ramus et al., 1999; Grabe and Low 2002)。近年の音楽と話し言葉のリズムの比較研究によると、言葉のリズムは等時性に依拠するものではないという主張(Patel 2007:122)や音節に依拠するという主張(Jakendoff 2009:198)も見られる。

このようにモーラ・タイミングの用語は時間と深い関係があるような印象を受ける。だが、古典ギリシャ語の韻文のリズムは長短音節の組合せの反復が主因であり、timing unit としてのモーラは韻文を発話する際の指標的な位置づけにすぎない。すなわち、Jakendoff (2009)が指摘するように音節が言葉のリズムの本質であるとすると、古典ギリシャ語のモーラの概念をモーラ・タイミングに転用したこと自体が矛盾を生み出すことになる。また、Bloch (1950)は日本語の自然発話に対して聴覚的印象の観点からモーラに言及しているが、長短音節との関わりの中でのモーラの位置づけという最も根幹となる議論は行われてこなかった。そこで本発表は、韻文が発話行為を前提とするテキストであることに着目して、古典ギリシャ語の韻文で想定されたモーラの概念に包含される timing unit の本質について古代ギリシャ語と同程度の歴史を有するとされる古代インドのヴェーダ語の韻文のリズム（紀元前 1500 年から紀元前 200 年とされる。詳細は Macdonell (1900)参照）を語彙認識の観点から比較検証を行うことを目指す。

２．構造主義言語学とモーラ

本節ではアクセントと話し言葉のリズムの両探求におけるモーラについて構造主義言語学者及びその後の研究者の主な言及について論じる。

2.1 アクセントとモーラの解釈

・西欧語ではアクセント価１＝音節１が原則。非西欧語ではアクセント価１また2＝音節１の事例が発見される(Trubtzkoy 1939/1969)。音節に関する２種類の類型的用語提唱。(syllable-counting vs. mora-counting; phonetic vs. phonemic syllables)

・アクセント価の分布を基に音節を２項対立的に分類し、アクセント価を担う単位としてモーラを転用。長短音節内のモーラ数はアクセント価に対応させ２と１とした。

・このアクセント価を担うモーラに対して時間的属性を巡り研究者間で言及が異なる。

言及派：Pike 1947; Hockett 1955

非言及派：Trubetzkoy 1939/1969; Trager 1947

転向派：McCawley (1968)は言及 -> MaCawley (1978)は非言及

・ 今日の生成音韻論では非言及で統一(Odden 2011)(モーラは抽象的な単位と解釈される)。

・モーラに関して構造主義言語学者の中でBloomfield (1933:110)は独特の解釈をする。音素を抽象的な位置づけとし、発話時の分節音の相対的時間を示す恣意的単位モーラを提唱。母音や子音の時間を示すのに one mora や one-half morae とした。

・Pike(1947)やHockett (1955)は長音節のとの関わりで言及したのに対して、Bloomfieldにはこの言及はない。

2.2 モーラ・タイミングの探求とモーラの解釈

・「日本語の発話は「スタッカート」を叩くような聴覚的印象を受ける」(Bloch 1950)。

・日本語の音節が均等な時間で発話されると推定。Allen (1973: 59)は“What the Greek ear actually measured was not “the syllables”...but ...time taken...を紹介。

・「日本語の発話は均等な時間を有するモーラの連続からなる」という説を主張(Ladefoged 1975)。日本語の自然発話は等時性を有するモーラが「鎖の環」の連続と解釈していたことを示唆。すなわち、リズム＝語彙認識と同一視している。

・だが、この解釈は近代言語学の最大の発見である「音声言語の語彙は音素からなる」という主張と矛盾するのではないか。

・すなわち、Bloch (1950)やLadefoged (1975)の説は音声言語の根本の仕組みである音素レベルを飛び越えたものであるが、この矛盾が指摘されることはなかった。

・「鎖の環」の連鎖のように聞こえるのは時間軸に沿って音素認識が行われた後にモーラレベルで分節の境界を認識するから。すなわち、等時性の時間の認識ではなく、分節としてのモーラ機能の認識の存在を示唆 (Cutler and Otake 2002)。

(a) CV 音節（短音節） : CV.CV.CV. … (.がモーラ境界を示す)

(b) CVC 音節(長音節) : CV.C.CV. …

(c) CVV 音節（長音節） : CV.V.CV …

・つまり、長音節が 2 モーラ、短音節が 1 モーラと解釈されるのはモーラの等時性の認識ではなく、音素認識の後にモーラ境界が認識されることを示唆（Otake et al., 1993; Cutler and Otake, 2002 参照）。
・Patel (2007:60)は言葉のリズムは等時性と無関係であるとし、Jackendoff (2009)は話し言葉のリズムは時間ではなく音節（長短音節）に依拠することに言及。だが、彼らはオンラインレベルでどのようなプロセスが起こっているかに言及していない。
・すなわち、Bloch (1950)が日本語の発話のリズムを等時のモーラの時間に着目したものはモーラの分節の境界と解釈すると、古典ギリシャ語の韻文のリズムにおける timing unit としてのモーラは分節のモーラ境界と解釈できるのではないか。

３．韻文のリズムとは

3.1 韻文とは

・言語文化の表現様式には散文と韻文が存在する。前者は意味基準で語彙選択を行うのに対して、後者は意味に加えて語彙内の音節の反復が見られるように語彙選択を行う。
・韻文は口承を前提とする言語表現様式である。分割されたスタンザ（行）の内部に配置された音節がリズムを生成する音声化のプロセスが伴うことが前提とされる。

3.2 韻文のリズムの決定要因

・韻文のリズムは当該言語の文化が決定した音節の認定によって決まる。
・音節の認定方法は２種類：(1)音節数と(2)分類された音節の組み合わせ。
(1)：仏語、伊語、中国語、日本語など、
(2)：古典ギリシャ語、古典ラテン語、英語、ドイツ語など。
・言語文化圏と音節の認定
(a) グレコ・ローマン文化圏（古典ギリシャ語、古典ラテン語）=> (2)（音節の長短）
(b) ゲルマン語文化圏（独語、英語、蘭語） => (2)（音節の強弱）
(c) ロマンス語文化圏（仏語、伊語、西語） => (1)
(d) 中国語文化圏（中国語、日本語、韓国語） => (1)
・音節の長短は母音の長短が弁別的であることが前提。 (a) => Yes (b) => No

3.3 古典ギリシャ語の韻文のリズムの認知過程

・古典ギリシャ語の母音体系は母音の長さが弁別的。

(1) CV 音節	=> 短音節	default	母音長認識
(2) CVV 音節	=> 長音節	by nature	母音長認識
(3) CVC 音節	=> 長音節	by position	音素認識

すなわち、音節長の決定は母音長と音素の２つの要因の認識で決定。
・古典ギリシャ語の韻文のリズム体系は長短音節の恣意的組合せによるので無限に想定可能であるが、特定の組合せを選好。＜長短　短長　長短短　… ＞　言語文化が決定。
・古典ギリシャ語の韻文の発話を聞いたなら、長短音節はどのように認識されるのか。

(a)古典ギリシャ語は自然言語であるので音素の連鎖として発話されるはずである。
(b)CV 音節は子音と母音を認識　＝＞ CV.
(c)CVV 音節は母音が CV 音節の母音より長いことを認識　＝＞ CVV. [VV ＞ V]
(d)CVC 音節は短母音の後に子音音素が連続することを認識。＝＞ CVC. [VCC => VC.C]
(e)CVV 音節と CVC 音節を長音節とするのは約束事に過ぎない。

・すなわち、長短音節の認定はあくまでも音素レベルの認識がベースになるので、長短音節が「鎖の環」のように認識されることは起こり得ない。
・長短音節の認定基準（by nature と by position）は最古の古典ギリシャ語文法書に記述されていることから、timing unit としてのモーラはリズムの本質ではなく、余剰的であることを示唆している。
・モーラは母音の長さが弁別的でかつ音節を長短に分類する慣習を持つ言語にのみ存在。
・モーラは母音長の弁別が前提であるため CVV ＞ CV が認識できれば mora の認識は可能。又、CVC は CV.C と認識すればよい。この仕組は古典ギリシャ語の長短音節の分類原理“by nature”と“by position”と同じ。従って、timing unit の概念よりも分節の原理の方がより合理性が高い
補足：Greene and Cushman (2012)によると、モーラの項目が初めて取り上げられているが、韻文における mora の位置づけの言及はなく、構造主義言語学の定義の紹介に終始。

3.4　その他の言語の韻文のリズムと音節

・中世ラテン語：母音長が音韻変化により非弁別的となると、ロマンス語系の言語が生まれ、音節数がリズムの要因となる。
・中国語文化圏：一貫して音節数による。＜中国語、韓国語、日本語＞
・日本語：11 世紀までに中国語との言語接触や音便などで音韻変化が起こると長短音節が生まれる。だが、音節を長短で認定する慣習は生まれず、長音節を２項要素に分離する分節が発生。書記体系も追従。長音節を含む語彙は国風文化重視から忌避され、韻文には非反映。このように日本語には長短音節が存在するものの中国語文化圏の慣習である音節数に留まる。言語文化的に日本語の韻文は古典ギリシャのものとは全く異なる。

4. ヴェーダ語の韻文のリズム

・ヴェーダ語はインドのサンスクリット語の古形とされる。時代的には、紀元前 15 世紀に遡るもの。紀元前 5 世紀の「アレキサンダー大王のインド侵攻」の時代には独自の言語文化が存在していた（MacDonell 1900:7)。
・ヴェーダ語の韻文は以下のような特徴がある(Macdonell 1916:436)。
(1) スタンザ数は４行（quarter）。
(2) スタンザ内は固有の音節数（8 音節、11 音節、12 音節、まれに 5 音節）が配置。
(3) 音節は長短に分類。その認定方法は古典ギリシャ語と同様である。
(4) スタンザは 4 又 5 音節の前半部(opening)に caesura(休止)を入れ、長短音節の交代

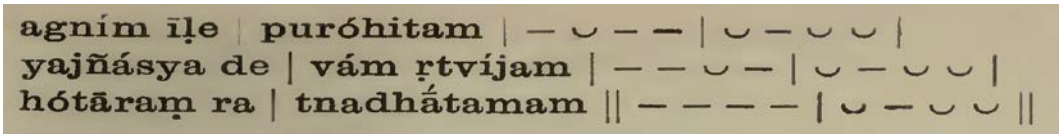
は後半部（cadence)のみに出現。

(5) スタンザの最初と最後の音節の長短の認定はなし。

(6) 長母音は２つの短音節と等価として発音せねばならない。

(6) 長短音節の認定は古典ギリシャ語と同様である。

(7) 例　The Gayatri stanza（二番目に一般的な形式、3 行 8 音節）
(Macdonell 1916 :438)



(a) 前半部(opening)は音節の長さは行内及び行間でパターン化されていない。

(b) 後半部(cadence)は iambic（短長音節の組合せ）の後に spondee（長長音節の組合が配置。

・前半部では長短音節の組合せを考慮せずに語彙配置が行われる。他方、後半部ではこの組合せが出現。各スタンザ内の音節の特定の形式が出現するのは部分的に過ぎない。一方、古典ギリシャ語では各スタンザ内での初めから最後までほとんどの部分に同一形式の反復が見られる点で大きな違いがある。従って、ヴェーダ語の韻文では timing unit を想定するまでもない。

・一方、古典ギリシャ語ではスタンザ内が一貫した音節の組合せが出現するために timing unit が存在するメリットはある。

5. 結論

・古典ギリシャ語の韻律のリズムは等時の timing unit としてのモーラで測定されると考えられてきた。だが、古典ギリシャ語は自然言語の一つであるので語彙認識の仕組みを前提としたリズムの認知システムとして説明する方がより普遍的な説明が可能となる。すなわち、古典ギリシャ語、ヴェーダ語、日本語の韻律のリズムについても同一の基盤で説明できる合理性がある。

・古典ギリシャ語は言語研究において文献上最古の言語として位置づけられてきた。そのため古典ギリシャ語で発生したモーラの概念は普遍的なものであるかのようにみなしてきたが、今後この考え自体再考の余地があるのかもしれない。

参考文献

Allen, S. W. *Accent and Rhythm*. Cambridge: Cambridge University Press.

Bloch, B. (1950) “Studies in colloquial Japanese II: syntax.” *Language* 26, 86-125.

Bloomfield, L. (1933) *Language*. New York: Holt, Rinehart and Winston.

Cutler, A. and T. Otake. (2002) “Rhythmic categories in spoken-word recognition.” *Journal of Memory and Language* 46(2), 296-322.

Dale, A.M. (1964) “Observations on Dactylic.” WS 77, 15.
Dauer R.M. (1983) “Stress-timing and syllable-timing reanalyzed.” *Journal of Phonetics* 11, 51-62.
Fox, A. (2000) *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. New York: Oxford University Press.
Grabe E. and E.L. Low (2002) Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner (Eds.) *Laboratory Phonology* 7, 515-546. Berlin, Germany: Mouton de Gruyter.
Hockett, C. (1955) *A Manual of Phonology*. Memoir II. Bloomington: IUP.
Jackendoff, Ray. (2009) “Parallels and nonparallels between language and music.” *Music Perception* 26(3), 195-204.
Ladefoged, P. (1975) *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
Macdonel, A. (1900) *A History of Sanskrit Literature*. New York: Appleton and Company.
Macdonel, A. (1916) *A Vedic Grammar for Students.* Oxford: The Clarrendon Press.
McCawley, J.D. (1968) *The Phonological Component of Japanese Grammar*. Mouton: The Hague.
McCawley, J.D. (1978) “What is a tone language?” Tone: *A linguistic Survey*. In Victoria Fromkin (ed.), 113-131. New York: Academic Press.
Odden, D. (2011) “The representation of vowel length.” *The Blackwell Companion to Phonology* 2, 1335-1361. Cambridge, MA: Blackwell.
Pike, K. (1947) *Phonemics: A Technique for Reducing Languages to Writing*. Ann Arbor:MUP.
Otake, T. (2015) “Mora and mora-timing.” In Haruo Kubozono (ed.) *Handbook of Japanese Phonetics and Phonology*. 493-523. Berlin, Germany: De Gruyter Mouton.
Otake, T., G. Hatano, A. Cutler and J. Mehler. (1993) Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32, 258-278.
大竹孝司 (2015) 「モーラの時間的属性の再考」『音韻研究』19, 59-66.
Patel, A.D. (2008) *Music, Language, and the Brain*. New York: Oxford University Press.
Ramus, F., M. Nespor and J. Mehler (1999) “Correlates of linguistic rhythm in the speech signal.” *Cognition* 73, 265-292.
Thrax, D. (1874) *The Grammar of Dionysios Thrax*. (Translated by Thomas Davidson). St. Louis, Mo.: Studley.
Trager, G. L. (1941) “The Theory of Accent System.” *Language Culture and Personality*. 131-145. Menasha, Wis.: Sapir Memorial Publishing Fund.
Trubetzkoy, N. (1939/1969) *Principles of Phonology*. (Translated by C.A.M. Baltaxe from *Grundzuge der Phonologie*). Berkeley: University of California Press.

C6

A preliminary acoustic study of tone in Dzongkha

Seunghun J. Lee(International Christian Univ., Univ. of Venda), Shigeto Kawahara(Keio Univ.),
Haruka Tada, Hanna Kaji(International Christian Univ.)
kawahara@icl.keio.ac.jp, {c181433k, c181130s, seunghun}@icu.ac.jp

1. Introduction

Dzongkha, the national language of Bhutan, is a language with a two-way tonal contrast: high (H) and low (L). This study reports a preliminary acoustic analysis of Dzongkha tones, whose nature has not been studied instrumentally. The current data comes from a single native speaker of Dzongkha; hence, our data should be taken to be preliminary. However, since there have not been any instrumental analyses on the phonetics of this language, aside from an impresionistic description by van Driem (2015), we would like to situate the current study as a first stepstone toward more systematic phonetic analyses of this language, which are on-going.

2. Method

Dzongkha has eight vowels, transcribed by van Driem (2015) as /a/, /ä/, /e/, /i/, /o/, /ö/, /u/, and /ü/. Each of these vowels were read by a native speaker of Dzongkha with H-tone and L-tone. Both f0 and spectral natures of these vowels were analyzed using Praat (Boersma 2001). In addition to these vowels read in isolation, 34 H-tone tokens and 33 L-tone tokens, varying in vowel quality and onset consonants, were also read by the same speaker. The f0 patterns of these syllables were analyzed. Finally, we addressed one consonant-tone interaction in Dzongkha by examining 18 syllables with a voiced onset and 16 syllables with what has been referred to as “devoiced” consonants (van Driem 2015).

3. Results

3.1. Vowel quality

We first started by exploring the acoustic nature of each vowels that exist in Dzongkha (/a/, /ä/, /e/, /i/, /o/, /u/, /ö/, /ü/). Figure 1 plots the standard F1 and F2 chart of these eight vowels, which shows that for those vowels without umlaut signs (/a/, /e/, /i/, /o/, /u/), their F1 and F2 distribute in expected F1-F2 regions. We also observe that umlauted versions have lower F1, but, more clearly, higher F2 compared to non-umlaut versions, which suggests that they are fronted versions of the corresponding non-umlaut vowels (i.e. umlaut represents frontness, as in German). In Figure 1, four types of vowels are clustered in the left-top region (/i/, /e/, /ö/, /ü/). In order to explore how these sounds are distinguished acoustically, we examined their F3, which is known to distinguish front unrounded vowels from front rounded vowels (Reetz & Jongman 2009). The results appear in Figure 2. As expected, F3 distinguishes unrounded front vowels (/e/, /i/) and rounded front vowels (/ö/, /ü/), in

that the latter group has much lower F3.

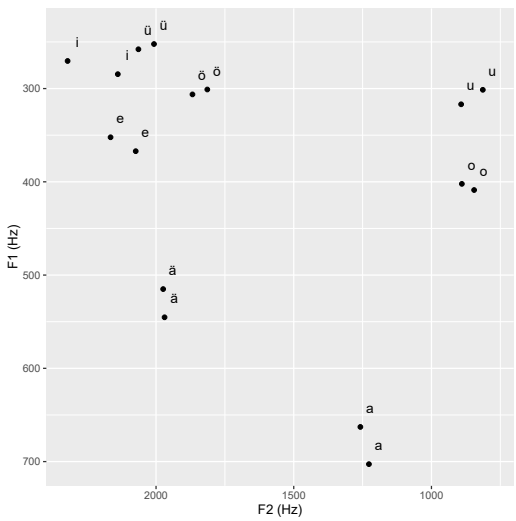


Figure 1: F1-F2 plot

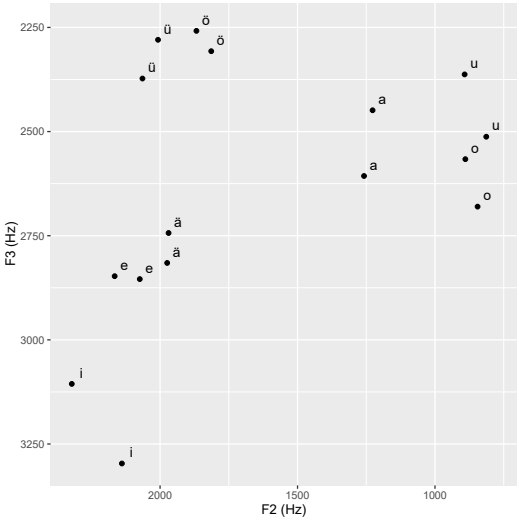


Figure 2: F2-F3 plot

3.2. Basic tonal patterns

Figure 3 shows raw f0 data of each vowel read in isolation. In this figure, all pitch points detected by Praat are plotted for the two tone types, presented by separate vowels. We generally observe separation between H-toned syllables and L-toned syllables in the first halves of the vowels. Some L-toned syllables (/ä/, /e/, /i/, /ö/, /ü/) show some bumps at their onset, to the degree that sometimes L-toned syllables have f0 as high as H-toned syllables. We are not confident whether these are simply measurement errors, or reflect something real (there is a generalization to be made here that only and all front vowels show this bump.) Besides these bumps, however, it is clear that H-toned syllables and L-toned syllables are separated in terms of their f0, especially at the beginning of syllables.

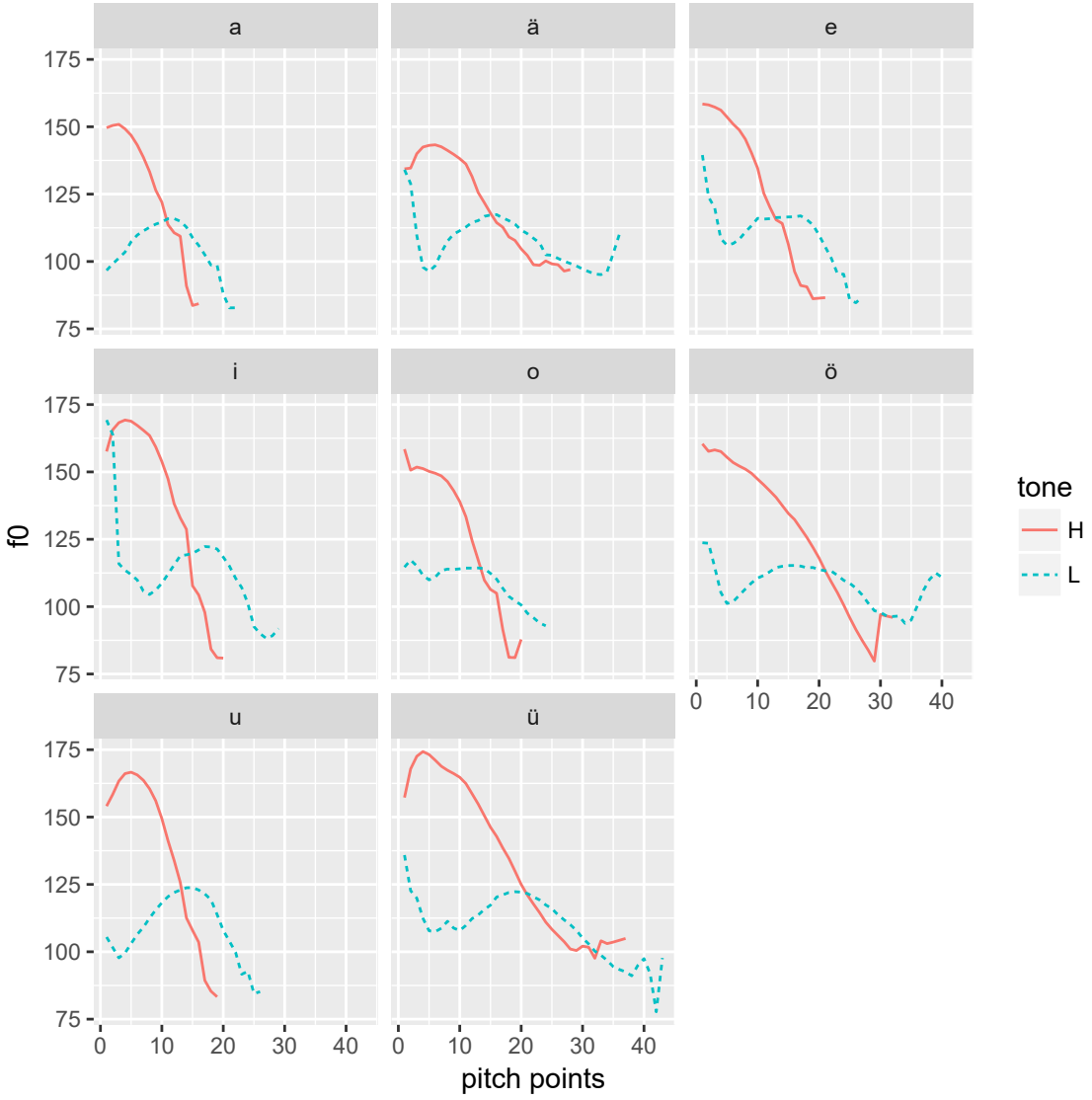


Figure 3: F0 movement of two types of tones by vowel (raw measurement values).

Figure 4 shows smoothed f0 curves which were obtained by deviding the vocalic intervals into five equally-timed windows, and taking the average f0 values within each window. It shows that H-toned and L-toned syllables are separated clearly at the onset of syllables, and the differences are neutralized toward the end for some vowels. The tonal difference seems to persist throughout the syllables for /a/, /i/, and /u/.

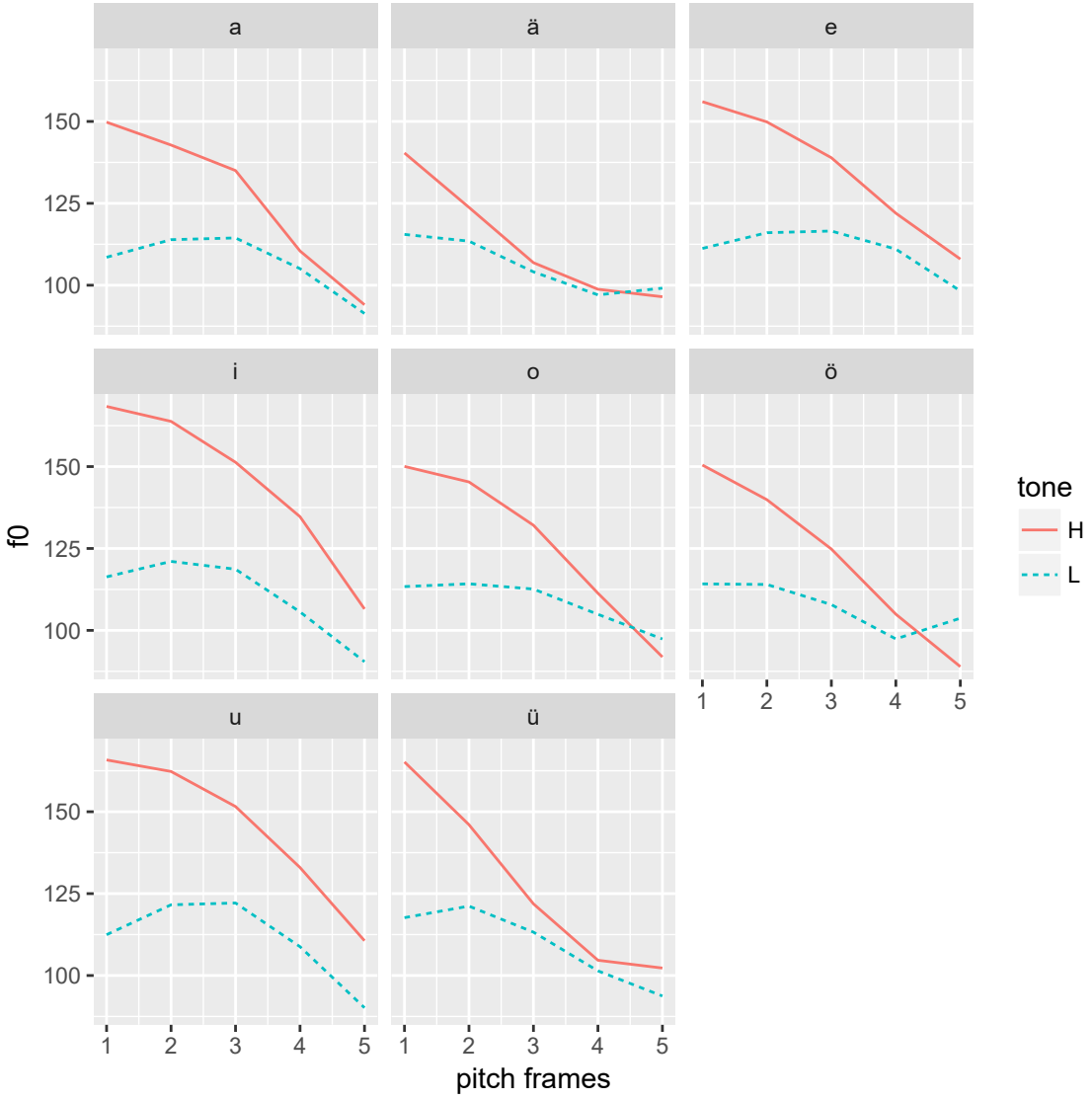


Figure 4: F0 movement of two types of tones by vowel (smoothed).

Figure 5 shows the average F0 plots of H-toned and L-toned syllables, based on tokens with onset consonants. Figure 5 is based on the smoothing analysis that is same as that of Figure 4; however, it targets only vocalic intervals. On average, at the onset of the syllables, H-toned and L-toned syllables differ by 30-40 Hz; the differences in F0 get smaller toward the end of the syllables, and are not observed in the final, fifth frame. What is emerging through our analysis is that tonal differences in Dzongkha manifest themselves at the onset of vowels.

In addition to the analysis of these f0 differences due to lexical H-tone vs. L-tone contrast, we also analyzed one type of tone-consonant interaction. Specifically, we examined 18 syllables with a voiced onset consonant and 16 syllables with what van Driem (2015) referred to as a "devoiced" onset consonant. The lexical tone of these syllables were L-tones. The result, which appears in

Figure 6, shows that the f0 is higher after voiced consonants than after “devoiced” consonants, the pattern that is opposite from what is expected it “devoiced” consonants were voiceless (e.g. Kingston & Diehl; Lee 2008).

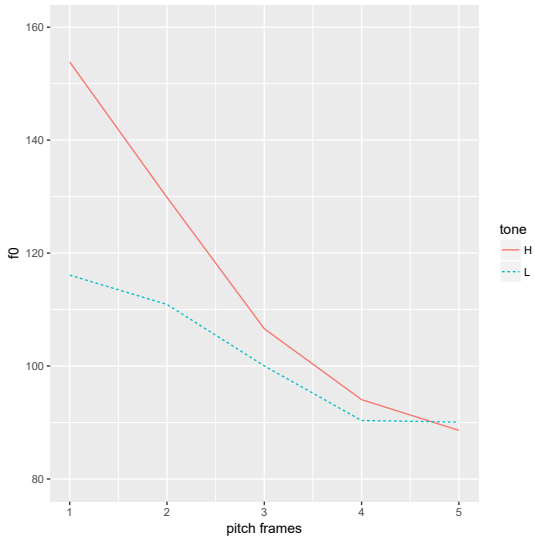


Figure 5: F0 differences of all syllables.

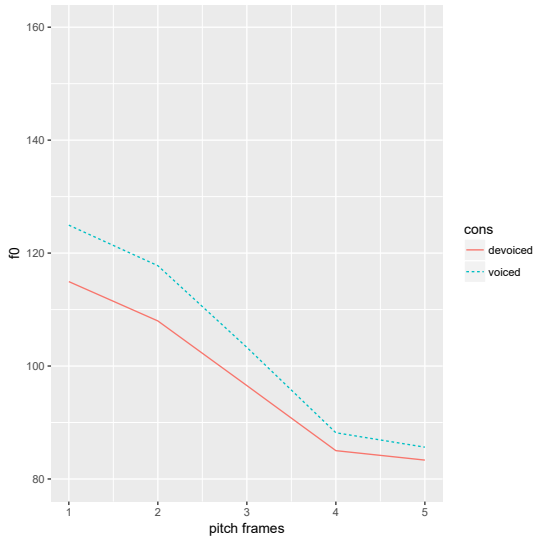


Fig. 6: Effects of “devoiced” consonants.

We thus entertain the possibility that the “devoiced” consonants are in fact breathy. In Bantu languages, breathy-voiced consonants lower the f0 of the following vowel (Baumbach 1987; Lee 2008). If these “devoiced” consonants were breathy, they follow the cross-linguistic pattern of pitch realization following breathy consonants. Although a more extended quantitative analysis is necessary, Figure 7 shows a spectrogram of “devoiced” [ba]. It shows that the consonant has long aspiration, which is compatible with the idea that “devoiced” consonants are in fact breathy.

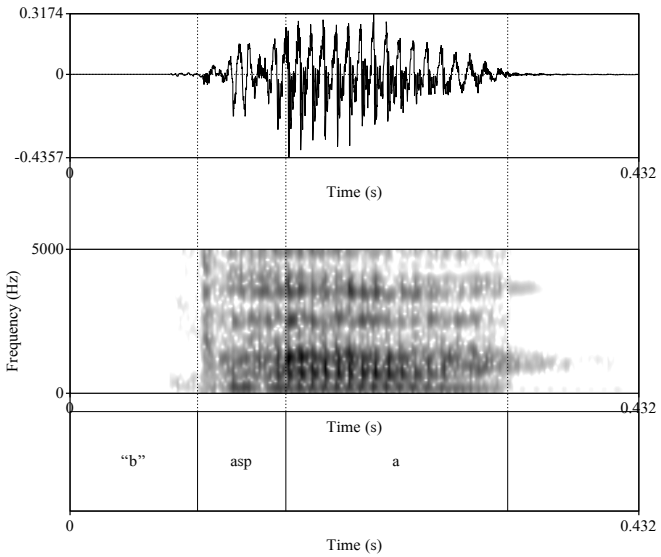


Figure 7: A waveform and spectrogram of "devoiced [ba]".

4. Conclusion

This paper has first examined the acoustic properties of eight vowels in Dzongkha. In addition to the "standard" /a/, /i/, /u/, /e/, /o/, the language has three vowels with umlaut which shows frontness. Those vowels with umlaut show higher F2 compared to the corresponding non-umlaut vowels. Rounding differences in front vowels manifest themselves in F3 in such a way that unrounded vowels have higher F3. Our f0 analysis shows that H-tones and L-tones in Dzongkha are distinguished at the beginning of vowels; i.e. the tonal targets are at the beginning. In some syllables, we observed cases in which f0 differences persist throughout the vowels. Finally, we showed that what has been referred to as "devoiced" vowels by van Driem (2015) raise f0 of the following vowels. We hypothesized that these "devoiced" vowels are in fact breathy. With this all said, our data is based on reading of a single native speaker; the analysis of more speakers of Dzongkha, and other related languages, is a topic of an on-going project.

References

Baumbach, E. (1987) "Analytical Tsonga Grammar." UNISA.

Boersma, P. (2001) "Praat, a system for doing phonetics by computer." Glot International 5: 341-345.

van Driem, G. (2015) "The Grammar of Dzhongkha." ms.

Kingston, J. & Diehl, R. (1994) "Phonetic knowledge". Language 70: 419-454.

Lee, S. J. (2008) "Consonant-Tone Interaction in Optimality Theory." Ph.D. dissertation. Rutgers University.

Reetz, H. & Jongman, A. (2009) "Phonetics: Transcription, Production, Acoustics and Perception". Blackwell-Wiley.

D1

How variations in syllable-final nasals affect word recognition in Taiwan Mandarin

Janice Fon, Yu-Ying Chuang (National Taiwan Univ.)
{jfon, d98142003}@ntu.edu.tw

1. Introduction

Human speech is full of variations. A simple word like *until*, the canonical (i.e., dictionary) form of which being /əntil/, can have a variety of pronunciations in daily conversation in addition to the prescribed form, including [əntʌ], [n̩tl̩], or even [tə] (Johnson, 2004). Therefore, how human listeners process spoken variants in their everyday lives is a topic that have intrigued researchers in the recent decade.

Interestingly, when facing the vast variability frequently encountered in speech production, human listeners do not seem to treat them in a uniform manner during spoken word processing. Instead, words in their canonical pronunciation seem to have an advantage and are processed more rapidly than words in non-canonical forms (e.g., Pitt, 2009; Ranbom & Connine, 2007; Sumner et al., 2014). Two accounts have been proposed for this processing bias. One explanation is to view human mental lexicon as being organized in a fashion similar to a dictionary, in which each lemma is phonologically coded in its canonical form only, and all its related variant forms are derived either via rules or associations (e.g., Gaskell & Marslen-Wilson, 1998; Gow, 2002). A canonical form thus enjoys processing efficiency not shared by its variant counterparts due to an exact mapping between the stored underlying form and the incoming speech signal. On the other hand, Sumner et al. (2014) argued that it is not the canonical status that makes a word form special, but rather, it is its social connotation that is granting it a distinctive place. They believe that the representation robustness of a word form is dependent on its social connotation. Word forms with positive connotation (i.e., more standard) are more strongly represented in the mental lexicon than words with negative connotation (i.e., less standard). In other words, the so-called "canonical form advantage" observed in many previous studies is in fact a "standard form advantage" instead, and the misconception comes from an alignment between canonical forms and standard forms that is frequently observed in languages.

This study thus intends to investigate the underlying cause for the spoken word processing bias using a particular set of syllable-final nasal mergers found in Taiwan Mandarin. This variation set is interesting because unlike most variations studied in previous research, the alignment between the canonical and the standard form is not always observed. There are in total three final nasal mergers, including /iŋ/→[in], /əŋ/→[ən], and /in/→[iŋ] (Fon et al., 2011). For /iŋ/→[in] (e.g., *gaoxing* 'happy' /kaʊɛiŋ/→[kaʊɛin]), the canonical [iŋ] is considered to be more standard than the variant [in] across dialects, showing perfect alignment. For /əŋ/→[ən] (e.g., *xueshen* 'student' /ɛɥeʂəŋ/→[ɛɥeʂən]), the canonical [əŋ] is considered to be more standard in some dialects, but the

variant [ən] is considered to be equally standard in others, showing only partial alignment. Finally, for /in/→[iŋ] (e.g., *aixin* 'a caring heart' /aɪɕin/→[aɪɕiŋ]), both the canonical [in] and the variant [iŋ] are considered to be equally standard across dialects, showing complete misalignment. In other words, one could consider the connotation of /iŋ/→[in] as mostly negative, that of /əŋ/→[ən] as relatively neutral, and that of /in/→[iŋ] as fairly positive. Therefore, if processing bias stems from differential social connotations of word forms [cf. Sumner et al. (2014)], then listeners should process the standard yet non-canonical [iŋ] from the /in/→[iŋ] merger, and, to a lesser extent, the standard-to-slightly-negative and non-canonical [ən] from the /əŋ/→[ən] merger with equal proficiency as their canonical counterparts of [in] and [əŋ], respectively. For the /iŋ/→[in] merger, listeners should still process the standard, canonical [iŋ] more efficiently than the nonstandard, non-canonical [in]. On the other hand, if processing bias is truly based on the canonical status of a word form [cf. Gaskell & Marslen-Wilson (1998); Gow (2002)], then listeners should only show a processing advantage for the canonical forms in all three mergers, regardless of its social connotations.

2. Method

2.1. Participants

A total of 198 native Taiwan Mandarin listeners, aged between 18 and 30, were recruited. Among them, 54 participated in Experiment 1, and 144 participated in Experiment 2, which had three sub-parts, one for each nasal merger, and each with 48 participants. None of the participants reported any language- or hearing-related disorders.

2.2. Materials

Experiment 1 included 45 bisyllabic stimuli, 15 for each nasal merger, and Experiment 2 included 90 bisyllabic stimuli, 30 for each nasal merger. These words were carefully selected so that both the canonical and the variant realization of the words are not homophonous with any other words in Mandarin. For example, both [kaʊɕiŋ] and [kaʊɕin] can only refer to *gaoxing* 'happy' /kaʊɕiŋ/, if at all. There were three prime conditions for each word, canonical, variant, and control. Canonical primes were realized with the dictionary form (e.g., [kaʊɕiŋ]), variant primes were realized with the appropriate variant form (e.g., [kaʊɕin]), and control primes were other bisyllabic words that were phonologically and semantically unrelated to the target words (e.g., *kafei* 'coffee'). For Experiment 2, there were also two target conditions for each word, canonical and variant. Table 1 shows an example of the prime and target conditions. For Experiment 1, all stimuli were recorded by the first author, who is a native speaker of Taiwan Mandarin. For Experiment 2, an additional male native speaker was also recruited to do the recording to avoid direct voice priming. Two lists were created via cross-splicing so that in each list, half of the trials were with male primes and female targets while the other half were with female primes and male targets. All stimuli were hand-adjusted to have the same duration using Cool Edit Pro (Version 2.0).

Table 1: An example of a stimulus *gaoxing* 'happy' /kaʊɕiŋ/ and its prime and target conditions.

	Prime (auditory)			Target		
	Canonical	Variant	Control			
Exp. 1	[kaʊɕiŋ]	[kaʊɕin]	*kafei* 'coffee'	高興		(visual)
Exp. 2	[kaʊɕiŋ]	[kaʊɕin]	*kafei* 'coffee'	Canonical:	[kaʊɕiŋ]	(auditory)
				Variant:	[kaʊɕin]	

2.3. Procedure

Both experiments were conducted in a sound-attenuated room. For Experiment 1, each trial began with an auditory prime, immediately followed by a visual target. For Experiment 2, an auditory prime was followed by an auditory target after a 500 ms pause. For both experiments, listeners were required to perform lexical decision on the visual/auditory targets as accurately and rapidly as possible. The experiment lasted about 20 and 40 minutes for Experiments 1 and 2, respectively, and listeners were compensated for their participation with a small monetary reward.

3. Results

Response time (RT) was measured from the onset of target presentation, and trials with incorrect responses (i.e., recognizing canonical or variant targets as nonwords) were excluded from RT analyses. For Experiment 1, a Prime (3) × Merger (3) ANOVA was performed (Figure 1). Results showed that only the main effects were significant [Prime: $F(2, 2372) = 116.33$, $p < .001$; Merger: $F(2, 2372) = 5.32$, $p < .01$]. Post hoc analyses indicated that control primes were always slower than the other two conditions ($p < .001$), but the difference between canonical and variant forms was not significant. RTs for /əŋ/→[ən] were overall longer than those for /iŋ/→[in] and /in/→[iŋ]. However, this was mainly contributed by the control primes, and exclusion of such removed the effect.

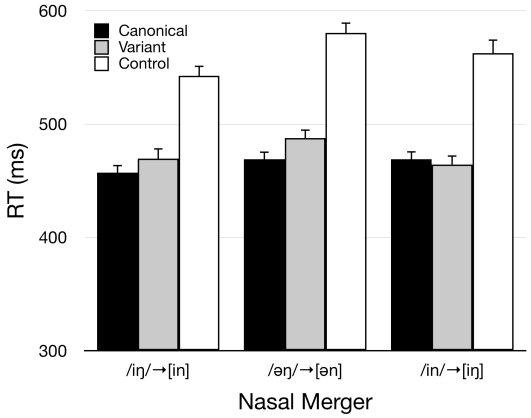


Figure 1: RTs of the three priming conditions in Experiment 1.

For Experiment 2, a Prime (3) × Target (2) × Merger (3) × Voice gender (2) ANOVA was performed. All main effects reached significance [Prime: $F(2, 8353) = 134.12$, $p < .001$; Target: $F(2, 8353) = 9.75$, $p < .05$; Merger: $F(2, 8353) = 19.56$, $p < .001$; Voice gender: $F(2, 8353) = 38.19$, $p < .001$]. Two interactions involving Target and Merger were also significant [Target × Merger: $F(2, 8353) = 18.67$, $p < .001$; Target × Merger × Voice gender: $F(2, 8353) = 3.09$, $p < .05$]. Post hoc analyses showed that the control primes were again with longer RTs than canonical and variant primes ($p < .001$), while the difference between the latter two was again not significant. For /iŋ/→[in], targets realized as variant forms (i.e., [in]) were recognized more slowly than those as canonical forms (i.e., [iŋ]) ($p<.001$) (Figure 2a). In contrast, for /in/→[iŋ], variant targets (i.e., [iŋ]) were more rapidly recognized than canonical targets (i.e., [in]) ($p < .005$) (Figure 2c). The patterns held true for both male and female voice. However, for /əŋ/→[ən], there was an additional interaction with voice gender. Variant targets (i.e., [ən]) were realized less efficiently than canonical ones (i.e., [əŋ]) only when they were uttered with the female voice ($p < .01$). The effect disappeared when the targets were in the male voice, and both canonical and variant targets were realized with comparable speed.

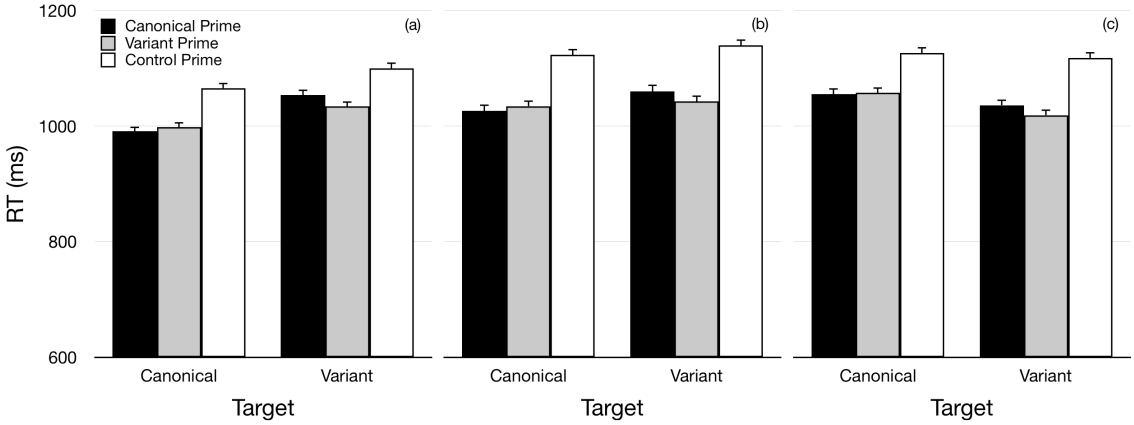


Figure 2: RTs for (a) /iŋ/→[in], (b) /əŋ/→[ən], and (c) /in/→[iŋ] in Experiment 2.

4. Discussion

This study examined the effect of a set of syllable-final nasal mergers on spoken word recognition in Taiwan Mandarin. Speech processing of canonical and variant forms of the three mergers, which differ in their social connotation, was compared using two form priming experiments. Results of Experiment 1 showed that the canonical form and its variant primed their linked written Chinese characters with comparable efficiency. This was rather surprising, as previous studies have shown otherwise [e.g., English: Ranbom & Connine (2007); French: Snoeren et al. (2008)]. Regardless of its origin, some kind of priming biases has always been observed. One suspects that this discrepancy might have been due to the peculiarity of the Chinese writing system, as its pronunciation encoding is not at all transparent. Thus, although written word recognition in alphabetic or syllabary writing inevitably involves phonological processing, this might not be necessarily so in Mandarin (Wu et al.,

2013). Compared to other languages, Mandarin users might be able to bypass detailed phonological processing when recognizing written word forms.

The uni-modal priming design in Experiment 2, on the other hand, revealed a significant effect of variation. Specifically, a robust effect of target realization was found on immediate spoken word recognition. Three different perceptual results were obtained. For /iŋ/→[in], the recognition of the canonical forms was more efficient than that of the variant forms. The processing advantage, however, greatly decreased for /əŋ/→[ən], since more efficient processing of the canonical forms only appeared in a voice-gender-dependent fashion. Critically, the result of /in/→[iŋ] was the most different, given that the processing of the variant forms outperformed that of the canonical forms. The current results were therefore in stark contrast to the hypothesis of the traditional account [cf. Gaskell & Marslen-Wilson (1998); Gow (2002)], because words encoded in canonical forms were not always more efficiently recognized than those in variant forms. They however strongly supported the social connotation hypothesis [cf. Sumner et al. (2014)], as the processing pattern was much in line with the connotation of the mergers. Based on the current findings, it can be seen that processing efficiency of a given form is not determined by whether it conforms to the canonical form. Instead, it is determined by whether this form is socially ideal or not. Previous findings of more robust processing for canonical forms thus likely stems from a frequent alignment between canonical forms and standard forms, rather than a true canonical form advantage in the mental lexicon.

However, similar to Experiment 1, priming difference between canonical and variant forms was still lacking in this experiment. This is surprising, as one would expect that same forms should prime better than different forms. However, as shown in Figure 2, one did observe a slight trend for a priming advantage between the same forms, and thus one suspects that the lack of significance might have come from the current experimental design. Given that there is a long interval of 500 ms between the offset of the primes and the onset of the targets, it is likely that any processing difference that might have existed between the two forms has already diminished or even disappeared by the time when auditory targets were presented. For future studies, inter-stimulus intervals could be manipulated in order to examine the time course of the priming effect in variant realizations.

5. Conclusion

By examining the three final nasal variations in Taiwan Mandarin, this study has thus demonstrated that it is social connotation that plays an important and decisive role in spoken word recognition, not canonical status. Socially ideal forms are more felicitous to lexical access than socially non-ideal ones, a finding much consistent with the socially weighted model proposed by Sumner et al. (2014). However, such an effect seems to be fairly short-lived and modality-sensitive. Unlike other languages [e.g., Ranbom & Connine (2007), Snoeren et al. (2008)], in which differential priming effects between canonical and variant forms have been consistently observed, even in cross-modal

experimental setups, no such evidence was found in the current study. Although one could probably attribute the null effect of differential priming from Experiment 1 to the logosyllabic nature of Chinese characters, the null effect of differential priming from the second experiment is not as easily explicable. Even though it is likely that the 500 ms inter-stimulus interval employed in this study is too long to detect a preferential priming effect, as mentioned above, Sumner & Samuel (2009) used exactly the same interval for their English participants and obtained significant results. It is thus possible that there might be cross-linguistic differences in spoken language processing with regards to the rate of dissipation in priming. It is also possible that different variations might have been the cause, as Sumner & Samuel (2009) examined syllable-final rhotics in English, and the current study looked at syllable-final nasals in Mandarin, which are acoustically less robust. More studies would be needed in order to understand the underlying cause for the preferential effects and the non-effects found in spoken variant processing.

References

Fon, J., Hung, J.-m., Huang, Y.-H., & Hsu, H.-j. (2011) "Dialectal variations on syllable-final nasal mergers in Taiwan Mandarin", Language and Linguistics 12(2), 273-311.

Gaskell, M. G. & Marslen-Wilson, W. D. (1998) "Mechanisms of phonological inference in speech perception", Journal of Experimental Psychology: Human Perception and Performance 24(2), 380-396.

Gow, D. W. (2002) "Does English coronal place assimilation create lexical ambiguity?", Journal of Experimental Psychology: Human Perception and Performance 28(1), 163-179.

Johnson, K. (2004) "Massive reduction in conversational American English." In K. Yoneyama and K. Maekawa (eds.) Spontaneous Speech: Data and Analysis. (pp. 29-54). Tokyo: The National International Institute for Japanese Language.

Pitt, M. A. (2009) "The strength and time course of lexical activation of pronunciation variants", Journal of Experimental Psychology: Human Perception and Performance 35(3), 896-910.

Ranbom, L. J., & Connine, C. M. (2007) "Lexical representation of phonological variation in spoken word recognition", Journal of Memory and Language 57(2), 273-298.

Snoeren, N. D., Seguí, J., & Hallé, P. A. (2008) "On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French", Cognition 108(2), 512-521.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014) "The socially weighted encoding of spoken words: a dual-route approach to speech perception", Frontiers in Psychology 4.

Sumner, M. & Samuel, A. G. (2009). "The effect of experience on the perception and representation of dialect variants", Journal of Memory and Language 60(4), 487-501.

Wu, J.-T., Yang, F.-L., & Lin, W.-C. (2013) "Beyond phonology matters in character recognition", Chinese Journal of Psychology 55(3), 289–318.

D2

東京語動詞の打消し形のアクセントの生成
—中国北方方言話者によるアクセントの特徴とその要因—

王　凰翔（広島大学大学院）
d164478@hiroshima-u.ac.jp

1. はじめに

中国語母語話者が対面コミュニケーションで日本語母語話者に話す際に、誤ったアクセントで発音すると、しばしば誤解を招くことがある。一方、日本語教育の面では、音声教育に関する教材が殆どないという状況があり、これまで行われた日本語検定でも、アクセントの問題は出されていない。このことは、日本語学習者に対する発音の指導が未だに重視されておらず、また、多くの学習者が日本語アクセントに対する知識を十分持っていないことを示す。そこで、筆者は動詞の打消し形に焦点を当てて、中国北方方言話者(上級日本語学習者)を対象として生成調査を行い、学習者の習得実態及び形成要因を明らかにする。

2. 先行研究及び先行研究の問題点

アクセント動詞の生成研究について、劉(2008)は中国語話者(北京・上海)を対象とし、動詞の辞書形、9 つの活用及び複合動詞アクセントの生成調査を行い、習得が容易な部分と習得が困難な部分、さらに誤用パターンの傾向を指摘した。劉(2008)の観察した結果のうち、動詞未然形(「ない」形)は、生成調査の結果では、正答率が全体的に低い。また、北京方言話者と上海方言話者の間に生成の誤用パターンの差異が見られたと述べている。

これにより、中国語母語話者(北京・上海)には、動詞の打消し形「ない」の習得が最も困難であること、また、方言差のあることが分かった。しかし、なぜそういう結果が出てくるのか。その要因については、まだ十分に考察されていない。

3. 生成調査

3.1. 調査対象

本研究は、方言の影響を少なくするため、中国北方方言話者（上級日本語学習者)30 名を対象とし、生成調査を実施した。(ただし、2 拍一段動詞の被調査者は 20 名であった)

3.2. 調査手順

生成調査：日本語学習者 30 名に調査用語を自然に読んでもらい、その発音を録音する。生成調査に要する時間は一人につき約 2 分である。

3.3. 調査用語

動詞活用形	起伏式五段動詞	平板式五段動詞	起伏式一段動詞	平板式一段動詞
打ち消し形	飲まない	呼ばない	出ない	寝ない
	頼まない	遊ばない/及ばない	食べない	真似ない
	喜ばない	働かない	別れない	忘れない

4. 生成調査—結果の分析と考察

動詞打ち消し形の正答率について、その結果は、以下のようになる。

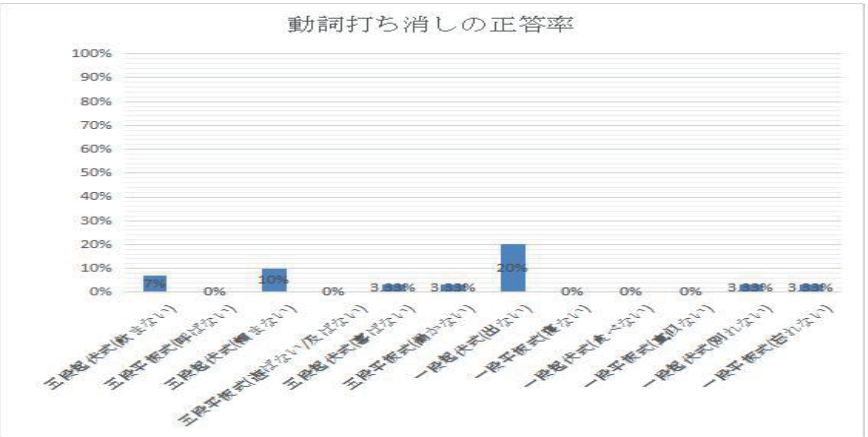


図 1: 動詞打ち消し形の正答率

図１より、打ち消し形（～ない）の正答率が極めて低く、学習者にとって打ち消し形の習得が最も困難であるということが分かる。平均正答率はわずか 7%である。しかも、そのうち、平板式動詞の正答率が極めて低いことが見てとれる。

表 1: -2 型アクセントで動詞打消しを発音する話者の比率(%)

動詞	飲まない	頼まない	喜ばない	出ない	食べない	別れない
-2 型	93%	90%	96%	80%	97%	97%
	呼ばない	遊ばない/及ばない	働かない	寝ない	真似ない	忘れない
-2 型	100%	90%	90%	100%	97%	97%

さらに、表１から、多くの学習者はすべての打ち消し形を-2 型アクセントで発音する傾向のあることが明らかになった。では、この結果が形成される要因はなんであろうか。筆者は下記のように四つの側面からこれを分析した。

4.1. 形容詞（〇〇ない）との関わり

まず、形容詞（〇〇ない）の習得が動詞の打ち消し形の習得に影響を与えるのではないかと考え、追加調査を実施した。調査者のうち、10 名の学習者に「汚い」「儚い」「拙い」「切ない」「危ない」の発音をしてもらった。その結果、全員が同じアクセント型で形容詞と打ち消し形を発音することが分かった。この結果から、形容詞（〇〇ない）の習得が動詞の打ち消し形の習得に影響を与える可能性が高いと言える。

表 2: 各教材の形容詞と動詞打ち消しの学習の順番

総合日本語	形容詞 A1　第 2 章－第 2 課	動詞打ち消し形　第 2 課－第 3 課
新編日本語	形容詞 A1　第 4 課	動詞打ち消し形　第 12 課
標準日本語	形容詞 A1　第 9 課	動詞打ち消し形　第 19 課
みんなの日本語	形容詞 A1　第 9 課	動詞打ち消し形　第 20 課

また、表2から分かるように、形容詞(○○ない)と動詞打消し形を学ぶ順番について、いずれの教材でも、日本語学習者は動詞打ち消し形より先に形容詞を学ぶようになっている。日本語学習者は先に形容詞のアクセントを習得するので、動詞打ち消し形を学ぶ際に形容詞と同様のアクセントパターンで読むという誤った先入観が生じやすい。かつ、日本語の形容詞のアクセントも動詞のアクセントと同様に、起伏式と平板式に分けられる。平板式のうち、2拍語はなく、3拍語以上も少数である。それに対して、起伏式は、拍数にかかわらず語末から2番目の拍にアクセント核が置かれることが一般的である。さらに、形容詞の終止形はアクセントのゆれがある。特に、平板式である語が多数派である起伏式で発音されることがよくある（加藤・安藤 2016:184)。現在は、東京方言でも若年層から中年層まで起伏式で発音する人が多くなっている（加藤・安藤 2016:184)。よって、日本語学習者が「ない」を含む形容詞を-2型アクセントで発音する確率は極めて高いと言える。

以上のことから、日本語学習者は「ない」を含む形容詞のアクセント（例：きたな’い、はかな’い)の影響を強く受けるため、動詞打ち消し形「～ない形」は形容詞アクセントパターンのように発音してしまう。つまり、形容詞と動詞打ち消し形の学習の順番の違いが、打消し形の正答率に影響を与える要因の一つであると言える。

4.2. プロミネンスとの関わり

二つ目の要因はプロミネンスに関わると考えられる。言語で何かを表現するとき、その中の特定の部分に音声的に焦点を当てて強調することがよくある。

われわれは普段感情をこめて否定文を発する際に常に打ち消ししている部分を強める傾向がある。ここで、いくつかの言語の否定文を見てみる。

例えば、「Q:あなたは林檎を食べたいですか？　A：(私は林檎を)食べたくないです。」という文があったとする。

1)日本語

Q:あなたは林檎を食べたいですか？

A：~~(私は林檎を)~~食べたくないです。

2)ドイツ語

Q：Möchtest du einen Apfel essen？

A: Nein, ich möchte nicht ~~(einen Apfel essen).~~

3)英語

Q：Do you want to eat an apple ？

A:No, I don’t want ~~(to eat an apple)~~.

4)中国語

Q:你想吃苹果吗？

A:不，我不想吃~~(苹果)。~~

各言語の否定文に対する回答文を見ると、いずれの言語でも、回答文の動詞の目的語が質問文と一致する場合、回答文における目的語が省略されることが可能である。そうすると、文中で最も強調される部分は否定を表す部分になる。上の例が示すように、日本語の「ない」、ドイツ語の“nicht”、英語の“don’t”、中国語の“不”という二重下線をつけた部分である。

では、曲線声調を持つ中国語を例にして、中国語母語話者は否定文を言う際に声調はどのように変化するのか。音声を可視化すると図2のような形で表される。

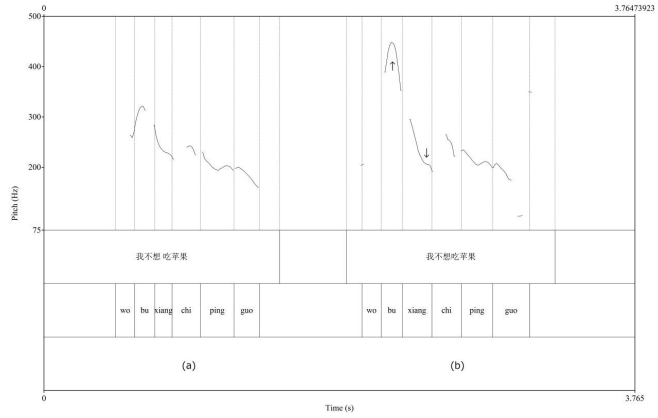


図 2: 我不想吃苹果[uo^{214} pu^{51} ɕiaŋ^{214} tʂʰi^{55} pʰiŋ^{35} kuo^{214}]

図 2 では、(a)は感情を込めず棒読みで「我不想吃苹果」を発音した。特にプロミネンスを置かない中立的な発話である。(b)は「不」のところにプロミネンスを置いた発話である。(a)の発話と比べ、(b)は「↑」で示した部分が高く、「↓」で示した部分は (a)より目立たなくなっている。つまり、プロミネンスを置くところでは高い部分がより高くなる。その語に続く部分の高さは目立たなくなる。

以上より、中国語における否定を表す字「不」は日本語における否定を表す「～ない」という部分に相当することで、中国語母語話者は日本語の動詞打消し形を発音する時、中国語の否定文を言うように、無意識的に打消しを表す部分を強調して発音してしまう。そうすると、日本語打ち消し形「ない」の「な」のところにアクセント核が置かれやすくなる。

4.3. 日本語の頭高型と中国語の第 4 声との関わり

三つ目の要因として、母語のアクセント[1]の干渉で発音し間違えることがある。中国語の第 4 声と日本語の頭高型のピッチ曲線を比較すると、中国語であれ、日本語であれ、ピッチ曲線は同じく高い位置から低い位置に急激に下降することが見られる。

従って、母語のアクセントの干渉で、中国語母語話者は日本語において打ち消しを表す部分「ない」を発音する時、母語の打ち消ししている字「不[pu^{51}]」のピッチ曲線のように発音してしまう。

4.4. 日本語の平板式と中国語の第 1 声との関わり

表 1 より、多くの学習者は-2 型アクセントですべての打ち消し形を発音することが分かる。この現象に関して、劉 (2008)、蔡(1983)では、中国語母語話者にとってピッチを平坦に保つことは困難であり、ピッチが下降する傾向があると述べている。しかしながら、このような傾向になる原因はなんであろうか。筆者は以下のように考える。

日本語の平板式アクセントに関しては、窪園(1998:84)は日本語（東京方言）では全体の語彙の約半数が平板型で発音され、そのために、下記例文(1)のような平板式の語だけからなる文も珍しくないと述べている。「例文(1)フランスの首相とブラジルの首相が明日アメリカで会談する」(窪園 1998:84)。

---

[1] ここでは、アクセントを広義の意味で用いる。

これに対して、中国語の名詞においては、3 音節以上の音節を発音する際に日本語の 4 音節「マユダマ」、5 音節の「イキモドリ」、6 音節の「マユツバモノ」などのように、下がり目がなく、平坦に発音する名詞は殆ど存在しない。名詞以外でも「开开心心[$k^hai^{55}k^hai^{55}ɕin^{55}ɕin^{55}$]、稀稀拉拉[$ɕi^{55}ɕi^{55}la^{55}la^{55}$]」のような副詞も少数である。

また、第 1 声の声調符号が付いている漢字は単独で発音されるとき、「媽」[ma^{55}]のように超高平となるが、いったん二音節のセットになった時、平坦に発音するのではなく、「音節の間に波の山を形成したり、波の上昇部分や下降部分を形成する傾向が強く見られる（平井 2012:102)。

従って、中国語母語話者はピッチを平坦に保つことが困難である。中国語母語話者においては、日本語の音節が長くなると、どこかに下がり目を置く傾向が高くなるのである。

4.5. 日本語教材におけるアクセントの解説との関わり

数多くの中国の国立や私立大学で使用されている教材を分析の対象として、動詞打ち消しのアクセントについての解説を見てみる。

(1)「総合日本語」における動詞打ち消しのアクセントの解説

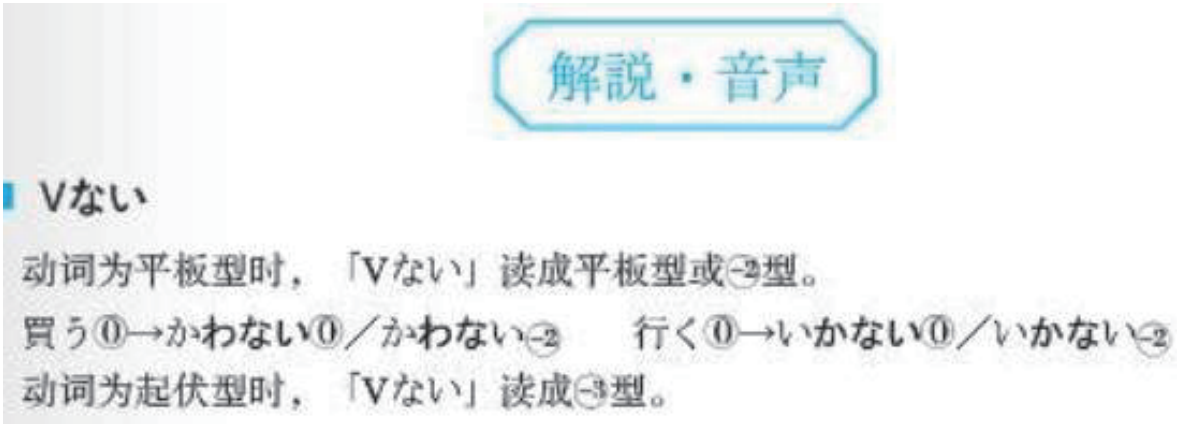


図 3:「総合日本語」における動詞打消しのアクセントの解説

上記の図 3 からわかるように、「総合日本語」という教材は、動詞打ち消しのアクセントに関して、起伏式動詞の場合は-3 型アクセントで発音され、平板式動詞の場合は 0 型または-2 型アクセントで発音されると述べている。

しかし、「総合日本語」における平板式動詞のアクセントの解説は受け入れがたい。東京語動詞の打ち消し形のアクセント型について、多くの言語学者(加藤・安藤(2016)，松森・新田・木部・中井(2012:35)，斉藤（1997:119))は、起伏式動詞の場合は-3 型アクセントで発音され、平板式動詞の場合は 0 型のみで発音されると述べている。しかし、「総合日本語」で述べられたアクセント法則はそれと一致していない。標準日本語（東京語、共通語、東京方言）に基づき、教材を作るのであれば、アクセントも標準日本語を基準として学習者に教えるべきであろう。「平板式動詞は 0 型または-2 型アクセントで発音される」ということを学習者に教えると、暗記しやすいために、平板式であれ、起伏式であれ、-2 型で

発音する人が多数になると考えられる。よって、今回の調査では、多くの人が-2 型で平板式動詞の打ち消し形を発音してしまったと考えられる。

(2)「新編日本語」における動詞打ち消しのアクセントの解説

「新編日本語」に関しては、「ない」の用法を解説する際に、アクセントの指導はされていない。単語帳には、名詞や動詞辞書形などはきちんとアクセント型がついているが、動詞と形容詞の活用形のアクセント型は解説していない。このことにより、学習者が動詞と形容詞の活用形のアクセント型に対する認識が不十分になることがうかがえる。

以上のことと生成調査の結果より、中国で出版された初級日本語教材における不適切で不十分な解説は日本語学習にマイナスの影響を与えると考えられる。不適切なアクセント型を学習者に教えると、学習者は覚えたままの形で、発音してしまうことが多い。初級日本語学習者は日本語の子音・母音（音素のレベル）やアクセント（かぶせ音素のレベル）を習得しないと、日本語を勉強すればするほど、無意識に自己流のアクセント型を形成する可能性が高い。これが、いったん癖になると、あとで正確なアクセントに直すのは簡単なことではない。

5. まとめ

本稿では、まず、中国北方方言話者を対象として動詞の打ち消し形の生成調査を実施し、学習者の習得実態を明らかにした。そして、筆者は動詞の打ち消し形と形容詞（○○ない）の学習順番、プロミネンス、両言語アクセントの相違、及び日本語教材におけるアクセントの解説という 4 つの側面から形成要因を考察した。また、本稿によって、学習者の習得実態を明らかにすると同時に、学習者自身の動詞活用形アクセントの問題に注意を喚起した。学習者に音声に対する意識を強く持たせることで、今後、日本語の学習に対してよりよい効果が得られるだろう。

参考文献

平井勝利（2012)『教師のための中国語音声学』白帝社

彭広陸・守屋三千代など（2009)『综合日语第一册』 北京大学出版社

加藤重広・安藤智子（2016)『音声学講義』研究社

窪園晴夫（1998)『音声学・音韻論』(西光義弘編「日英語対照による英語学演習シリーズ」1)くろしお出版

劉佳琦（2008)「東京語の動詞・複合動詞アクセントの生成について」『2006 年清華大学日本言語文化国際フォーラム論文集』清華大学出版社 pp. 424-439

松森晶子・新田哲夫・木部暢子・中井幸比古(2012):『日本語アクセント入門』研究社

斎藤純男（1997)『日本語音声学入門』三省堂

蔡全勝（1983)「中国人に見られる日本語アクセントの傾向」『在中華人民共和国日本語研修センター紀要　日本語教育研究論纂』第 1 号 pp. 26-31

周平・陈小芬（2009)『新编日语 第一册』上海外语教育出版社

D3

自己モニターを伴う日本語アクセントの産出訓練の効果
―中国語母語話者を対象として―

王 睿来（神戸大学大学院）・林 良子（神戸大学）
磯村 一弘（国際交流基金）・新井 潤（国際交流基金）
wang_ruilai@yahoo.co.jp

1. はじめに

日本語は語彙的なピッチアクセントを持つ言語であり、その習得は知覚面においても、産出面においても、多くの日本語学習者にとって困難であることがしばしば指摘される。また、それぞれの単語のアクセント型は恣意的に決まっているため、学習者はアクセントを単語ごとに覚えるしかないという性質を持っている。そのため、日本語アクセントの習得については、知覚と産出だけでなく、単語それぞれのアクセント型に関する知識も考慮する必要がある（王・林・磯村他，2017，p.62）。日本語アクセントの知覚・産出・知識の中で、中国語母語話者においては、その知覚の正答率よりも産出の正用率と知識の正答率のほうが低いことを筆者らはこれまでに報告してきた（王・林・磯村他，2017）。本稿においては、今回特に産出に着目し、中国語母語話者を対象に、産出訓練を行ない、その効果を検討した結果を報告する。

2. 先行研究

中国語母語話者による日本語名詞アクセントの産出に関する研究には、雷（2007）、柳（2010）、陳（2015）、王（2016）、王（2017）、王・林・磯村他（2017）などがある。そのうち、雷（2007）、柳（2010）と王（2017）は複合名詞の産出に関するものであるため、ここでは単純名詞を対象とした陳（2015）、王（2016）、王・林・磯村他（2017）について紹介する。

陳（2015）、王・林・磯村他（2017）では、中国語母語話者による日本語単純名詞アクセントの知覚・産出・知識の関係について検討した。その結果、産出の正用率が低く、知識が産出に影響を与えることが分かった。王（2016）では、学習者がアクセントをうまく発音できない場合、それは単語のアクセント型を知らないという知識の問題なのか、それともアクセント型を知っていたとしてもうまく発音できないという実現の問題なのかについて検討するために、3 つのタスクを用い、実験を行なった。タスク 1 では実験対象語を含めた短文の文字情報だけを、タスク 2 では文字情報にアクセント記号を示し、タスク 3 では文字情報にアクセント記号を示すと同時にモデル音声を呈示し、実験協力者に発話してもらった。その結果、アクセントの正用率は「タスク 1＜タスク 2＜タスク 3」のように有意に高くなった。この結果から、学習者がアクセントをうまく産出できないのは知識の問題だけでなく、実際にそれを発音する際の実現の問題もあることが分かった。また、モデル音声が正確な発音の実現を促すことが示唆された。

以上のことを踏まえ、本稿では日本語アクセントの実現を促すために、モデル音声を用いた産出訓練を行ない、その効果を検証した。

3. 方法

3.1. 実験協力者

実験協力者（以下、協力者）は中国の大学で日本語を主専攻とする 1 年生 24 名であった。全員が北方方言話者で、日本語学習歴は 6 か月程度であった。日本滞在経験がある者はなかった。24 名を 2 つのグループ（1）産出訓練を受ける産出訓練グループ（12 名）、(2) テストのみを受けるコントロールグループ（12 名）に分けた。

3.2. 実験対象語

実験対象語は 1 拍～4 拍の各アクセント型の単語で、合計 168 語を使用した。168 語中 84 語を訓練中に呈示し、残りの 84 語はテストのみで呈示した。本稿では前者を訓練用単語、後者を般化テスト用単語と呼ぶ。訓練用単語は「ま」「た」「ば」をそれぞれ拍数分用いて作った無意味語 42 語と協力者の未知語 42 語であった。般化テスト用単語はすべて協力者の未知語であった。

表 1: 実験対象語

訓練用単語		般化テスト用単語	
無意味語	未知語 1	未知語 2	未知語 3
42 語	42 語	42 語	42 語

3.3. 手続き

実験に入る前に、日本語のアクセントの仕組みについて解説を行なった。解説では、アクセントとは何か、アクセントの型、アクセントの制約、アクセントの記号について説明した。実験は pre-test、訓練、post-test の 3 つのフェーズで構成された。Pre-test と post-test では共通のテストを行なった。

表 2: 実験の流れ

フェーズ	グループ	
	産出訓練	コントロール
1	pre-test	pre-test
2	産出訓練	（訓練なし）
3	post-test	post-test

3.4. 産出訓練

産出訓練では、まず、平仮名で書かれた訓練用単語を助詞「が」付きでアクセント型とともに呈示し、自力で読み上げてもらい、その後、モデル音声を呈示し復唱してもらった。さらに、自力で発音したアクセントとモデル音声のアクセントが一致するかについて自己

評価し、プリントに「○」「×」で記入してもらった。自己評価させた理由は、モデル発音を繰り返し聞き、発音しながら形成した妥当な基準に基づいて自己評価する「モデル基準による自己モニター」が、発音能力に効果的な影響を及ぼす（小河原，1998）と考えられるからである。単語とモデル音声は PowerPoint を用い、パソコン画面で呈示した。呈示画面の例は図 1 のとおりである。モデル音声は東京方言を母方言とする日本人教員 4 名（男女 2 名ずつ）に録音してもらったものを使用した。1 つの訓練ブロックは訓練用単語 84 語をランダムな順に呈示した。2 訓練ブロック、合計 168 語を 1 回の訓練とし、2 名（男女 1 名ずつ）のモデル音声を使用した。1 回の訓練時間はおよそ 30 分で、8 回の訓練を 8 日に分けて実施した。

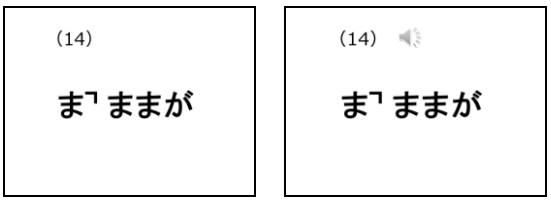


図 1: 産出訓練の呈示画面の例

3.5. テスト

Pre-test と post-test では共通の 3 つのテストを行なった。Test1 では、訓練用単語の未知語 1（表 1）を助詞「が」付きでアクセント型とともに呈示し、読み上げさせた。Test 2 では、般化テスト用単語の未知語 2（表 1）を用い、test1 と同じように読み上げさせた。Test 3 では、般化テスト用単語の未知語 3（表 1）を含めた短文をアクセント型とともに呈示し、読み上げさせた。Test1 と test2 では単語をすべて平仮名で呈示し、test3 では漢字がある場合は漢字で呈示した。テストは test1 から test 3 の順で行なった。3 つのテストを終了するのに、15 分～20 分程度要した。なお、本稿では test 3 を分析対象外とする。

4. 結果と考察

まず、全体の結果を示す。産出訓練グループとコントロールグループによる pre-test と post-test のアクセント正用率は表 3 のとおりである。コントロールグループによる pre-test と post-test の 4 テストはすべて 60%程度であった。その一方、産出訓練グループでは、pre-test の 2 テストはともに 50%程度であったのに対して、post-test の 2 テストはともに 100%近くであった。統計的に産出訓練の効果を検証するため、被験者間要因（グループ：産出訓練グループとコントロールグループ）と被験者内要因（テスト：pre-test の test1・test2 と post-test の test1・test2）という 2 要因の分散分析を行なった。その結果、交互作用も（$F(3,66) = 25.1$, $p < .001$）、単純主効果も有意であった。単純主効果について、グループの単純主効果は post-test の test1 と test2 において有意で（ともに $p < .001$）、テストの単純主効果は産出訓練グループにおいて有意（$p < .001$）であった。さらにテストの単純主効果について多重比較

を行なったところ、「pre-test の test1 < post-test の test1（p<.001）」「pre-test の test2 < post-test の test2（p<.001）」において有意であった。この結果から、コントロールグループでは、pre-test と post-test において、アクセント正用率に有意な変化が見られなかったが、産出訓練グループでは、post-test の test1 と test2 はそれぞれ pre-test の test1 と test2 より有意に上昇した。つまり、産出訓練は効果があったことが示唆された。また、訓練しなかった語に般化したと言える。

表 3: グループごとの pre-test と post-test の正用率（%）

テスト	pre-test		post-test	
	test1	test 2	test1	test 2
産出訓練グループ	53.6	57.7	98.8	97.4
コントロールグループ	61.1	62.1	62.9	66.1

つぎに、産出訓練グループの協力者は自己評価ができていたかについて検討する。そのため、産出訓練の最終回（8 回目）の自己評価で「自力で発音したアクセントとモデル音声のアクセントが一致する」と評価した割合を算出した。その結果、12 名による平均が 99.4%であった。これは post-test の test1 と test 2 の正用率とほぼ同じ程度であった。また、正しく自己評価するために、アクセントを正しく知覚する能力が必要であると思われる。この点を検証するために、訓練終了後、訓練用単語の無意味語 42 語を用い、知覚テストを行なった。その結果、全体の正答率は 92.7%と高かった。以上の結果から、協力者は高いアクセント知覚能力を持っており、自己評価が適切に行われた可能性が高いと言えるであろう。

さらに、産出訓練グループによる pre-test と post-test のアクセントの正用率と誤用アクセント型について検討する。上記の分散分析の結果、pre-test の test1 と test2 の間、post-test の test1 と test2 の間に正用率の有意差（それぞれ p=.37，p=.76）が検出されなかった。そのため、以下では、テスト別ではなく、pre-test の test1 と test2 を pre-test、post-test の test1 と test2 を post-test として、それぞれまとめて正用率（表 4）と誤用アクセント型（表 5）を検討する。

全体の正用率は、pre-test で 55.7%、post-test で 98.5%であった。拍数別の正用率について、pre-test でも post-test でも拍数が多くなるにつれて下がっていたが、pre-test より post-test のほうが下がり幅が小さく、どの拍数の正用率も 95%以上であった。この結果から、アクセントは拍数が多くなるにつれて産出が難しくなることと、産出訓練を通して、拍数が多くなっても正しく産出することが可能であることが示唆された。アクセント型別の正用率について、pre-test ではどの拍数においても 0 型の正用率が比較的高く、3 拍語の 2 型（40.3%）、4 拍語の 1 型（36.1%）、2 型（19.4%）、4 型（34.7%）が特に低かった。Post-test ではどの拍数のどのアクセント型の正用率も 90%以上であった。この結果から、産出訓練を通して、産出が難しいアクセント型であっても、正しく産出することが可能であることが示唆された。

表 4:産出訓練グループによる pre-test と post-test の正答数と正答率(%)

拍数	アクセント型	語数	pre-test (%)	post-test (%)
1 拍	0 型	72	47 (65.3)	72 (100.0)
	1 型	72	50 (69.4)	72 (100.0)
	合計	144	97 (67.4)	144 (100.0)
2 拍	0 型	72	62 (86.1)	72 (100.0)
	1 型	72	41 (56.9)	70 (97.2)
	2 型	72	39 (54.2)	72 (100.0)
	合計	216	142 (65.7)	314 (99.4)
3 拍	0 型	72	56 (77.8)	72 (100.0)
	1 型	72	39 (54.2)	70 (97.2)
	2 型	72	29 (40.3)	69 (95.8)
	3 型	72	44 (61.1)	72 (100.0)
	合計	288	168 (58.3)	283 (98.3)
4 拍	0 型	72	53 (73.6)	71 (98.6)
	1 型	72	26 (36.1)	67 (93.1)
	2 型	72	14 (19.4)	72 (100.0)
	3 型	72	36 (50.0)	72 (100.0)
	4 型	72	25 (34.7)	70 (97.2)
	合計	360	154 (42.8)	349 (96.9)
	全体の合計	1008	561 (55.7)	989 (98.5)

各拍のアクセント型別の語数＝実験対象語 3 語×協力者 12 名×2 テスト＝72 語

表 5:産出訓練グループによる pre-test の誤用数と誤用率(%)

拍数	ア型	0 型(%)	1 型(%)	2 型(%)	3 型(%)	4 型(%)	下げ上げ型[1](%)	合計(%)
1 拍	0 型	—	25(100.0)	—	—	—	—	25(100.0)
	1 型	22(100.0)	—	—	—	—	—	22(100.0)
2 拍	0 型	—	7(70.0)	3(30.0)	—	—	—	10(100.0)
	1 型	23(74.2)	—	8(25.8)	—	—	—	31(100.0)
	2 型	17(51.5)	16(48.5)	—	—	—	—	33(100.0)
3 拍	0 型	—	0(0.0)	1(6.3)	15(93.7)	—	—	16(100.0)
	1 型	21(63.6)	—	4(12.1)	7(21.2)	—	1(3.1)	33(100.0)
	2 型	23(53.5)	2(4.7)	—	18(41.8)	—	—	43(100.0)
	3 型	23(82.1)	1(3.6)	4(14.3)	—	—	—	28(100.0)
4 拍	0 型	—	0(0.0)	3(15.8)	7(36.8)	8(42.1)	1(5.3)	19(100.0)
	1 型	17(37.0)	—	0(0.0)	4(8.7)	15(32.6)	10(21.7)	46(100.0)
	2 型	17(29.3)	3(5.2)	—	19(32.8)	10(17.2)	9(15.5)	58(100.0)
	3 型	17(47.2)	0(0.0)	3(8.3)	—	14(38.9)	2(5.6)	36(100.0)
	4 型	24(51.1)	0(0.0)	3(6.4)	13(27.6)	—	7(14.9)	47(100.0)
	合計	204(45.6)	54(12.1)	29(6.5)	83(18.6)	47(10.5)	30(6.7)	447(100.0)

「ア型」＝「アクセント型」、「—」＝「該当なし」

誤用アクセント型について、pre-test（表 5）では全体の誤用傾向は、0 型ではないのに 0 型とする誤用が多く、全誤用数の 45.6%を占めた。これは王 (2016) と王・林・磯村他 (2017) での結果と一致する。アクセント別に見ていくと、1 拍語では、0 型を 1 型とする誤用と 1

[1] 「下げ上げ型」とは、ピッチが下がってからまた上がる誤用アクセント型のことである。

型を 0 型とする誤用がほぼ同数であった。2 拍語では、0 型を 1 型とする誤用と 1 型を 0 型とする誤用が多かった。2 型について、0 型とする誤用と 1 型とする誤用がほぼ同数であった。3 拍語では、0 型を 3 型とする誤用が多く、1 型・2 型・3 型について 0 型とする誤用が多かった。4 拍語では、0 型・1 型・2 型は多数の誤用アクセント型に分散していたが、3 型と 4 型について、0 型とする誤用が多かった。また、下げ上げ型の誤用について、3 拍語では 1 語、4 拍語では 29 語が観察された。Post-test では、4 拍語の 2 型では誤用が比較的多かった（全 19 語中 8 語）のに対して、他のアクセント型の誤用はすべて 2 語以下と少なく、下げ上げ型の誤用も観察されなかった。以上の結果から、産出訓練を通して、どんな誤用でもおおよそ修正することが可能であることが示唆された。

5. おわりに

本稿では、中国語母語話者の日本語学習者を対象に、自己モニターを伴う日本語アクセントの産出訓練を行なった。このことにより、このような産出訓練は大きな効果があること、非訓練語にも般化することを示す結果が得られた。また、訓練により、学習者は自分のアクセントの産出に対して、自己評価が適切に行われるようになる可能性も併せて示された。

謝辞

本稿は、文部科学省科学研究費：基盤（B）「海外日本語学習者音声アーカイブの構築・分析と WEB 韻律学習支援ツール開発」（課題番号：17H02352）による成果の一部である。

参考文献

小河原義朗（1998）『外国人日本語学習者の発音学習における自己モニターの研究』（博士論文，東北大学大学院文学研究科）

王睿来（2016）「中国語母語話者の日本語名詞アクセントの産出：アクセント核の情報とモデル音声の有無による影響に着目して」『第 11 回国際日本語教育・日本研究シンポジウム予稿集原稿』.

王睿来（2017）「中国語母語話者による日本語複合名詞アクセント産出：学習歴と単純名詞アクセント産出の影響」『ことばの科学研究』18，31-49.

王睿来・林良子・磯村一弘・新井潤（2017）「中国語母語話者による日本語名詞アクセントの習得：知識・産出・知覚の関係から」『中国語話者のための日本語教育研究』7，61-75.

雷寶茵（2007）「香港人広東語母語話者の複合名詞アクセントの生成について」『「日本語教育と音声」研究会第 7 回：創立 125 周年記念特別シンポジウム予稿集原稿』.

柳悦（2010）『中国人日本語学習者の複合名詞アクセント習得の縦断的研究：知識・発音・知覚の比較を中心に』（博士論文，首都大学東京大学院人文科学研究科）

陳冠霖（2015）「台湾人日本語学習者の 2 モーラアクセントについて：発音、聞き取り、記憶を中心に」『第 29 回日本音声学会全国大会予稿集』，24-29.

D4

L2 英語音声の評定における評定者の母語の影響

小西 隆之（早稲田大学大学院）近藤 眞理子 (早稲田大学)
tkonishi@aoni.waseda.jp, mkondo@waseda.jp

1. 概要

本研究では、第二言語 (L2) 英語音声コーパス (J-AESOP) に収録されている 183 名の日本人英語学習者による *The North Wind and the Sun* の読み上げ音声を計 5 人の英語母語話者および非母語話者が「セグメントの正確さ」などの 4 項目に関して評定したものの結果を比較分析し、異なる母語を持つ評定者間の評定値一致率や、評定が一致しやすい項目を明らかにした。

2. 本研究の目的

第二言語音声習得研究は、習熟度の異なる話者の比較分析を行い、その共通性と習熟段階における変化を検証するのが一般的である。この際、従来研究においては被験者募集のときに海外滞在年数などの明確な指標を用いて初級群と上級群を区分し、その二群の比較分析を行うという手法が用いられてきた。

しかしながら近年の L2 コーパスの普及により、中級群も含めた多様な学習者群をデータとして解析することが可能となった。L2 学習者は言語背景、教育環境や習得適性(aptitude)等の個人内要因も影響し、極めて多様な集団を形成する。特に L2 音声習得においてはこの多様性が顕著であると言える。

このような多様な集団を対象とした習熟度別の比較を行うためにはコーパスに評定値を付与する必要がある。Saito et al. (2015) はフランス語母語話者の L2 英語に 7 項目の評定値(「個々の子音や母音の発音は適切か」、「アクセント、イントネーションは適切か」等)を付与し、音声学者と非音声学者各 10 名の評定値の一致率を検証した結果、全ての項目について高い一致率を得た(クロンバックの α > 0.9)。

昨今では英語母語話者よりも L2 英語話者人口の方が圧倒的に多く(Crystal, 2003; Bolton, 2004 等)、英語母語話者が判断する理解容易度(intelligibility)だけでなく、L2 英語話者、とりわけ異なる言語背景を持つ L2 英語話者の理解容易度も考慮する必要がある。

そこで、本研究は、日本語母語話者の L2 英語音声の評定を対象とし、L2 音声の評定における評定者の母語の影響を検証することにより、L2 音声習得研究における評定基準の指標となることを目的とする。

3. 分析

評定対象としたのは J-AESOP に収録されている様々な習熟度の日本人英語学習者 183 名のデータで、評定対象とした音声は PC スクリーンに呈示された「The North Wind and the Sun」を被験者が音読したものである。評定者の個人内誤差を解消することと、特定箇所の

読み間違いなどが収録音声全体の印象を左右することを避けるのを目的として、各々の音声を 3 分割した。従って、評定作業に用いられたデータは 3(ファイル) x 183(名)の計 549 ファイルが評定者ごとにランダマイズされたものである。

本研究は英語母語話者 4 名、日本語母語話者 4 名、その他の言語の話者 8 名の総計 16 名に評定を依頼しているが、今回は既に分析作業が終わっている 1)英語母語話者 2 名、2)日本語母語話者 2 名、3)スペイン語話者 1 名による評定値をデータとして用いた。評定者はいずれも音声学を専門とする大学院博士後期課程の学生または大学教員である。

評定項目は a)個々のセグメントの正確さ、b)韻律の正確さ、c)流暢さ、d)母語訛りの少なさの 4 項目である(Table 1)。評定尺度は 1 から 10 までの 10 段階[1]で、数字が大きいほど評価が高い。

Table 1 評定者に呈示された評定項目(日本語訳)

項目	詳細
セグメントの正確さ (Segmental accuracy)	- 個々の母音や子音 (調音位置と調音法、有声無声対立など)
韻律 (Prosody)	- ストレスアクセント - リズム - 母音挿入やセグメントの脱落がないこと
流暢さ (Fluency)	- イントネーション[2] - 話速 - ポーズ
母語訛りの少なさ (Native-likeness)	- 母語訛りが少ないこと

4. 結果と考察

Figure 1 は評定値に関してそれぞれの評定者間で行った多重相関分析の結果である。EN は英語母語話者、JP は日本語母語話者、OT はその他の言語の母語話者であるスペイン語母語話者、01 と 02 は評定者 ID を表す。各セル内に相関係数が表示されており、セルの色の濃さは相関の強さを表す。全体を見ると、4 項目全てにおいて総じて強い正の相関が得られていることがわかる。

まず各項目内の評定者間相関を見ると、母語を共有する評定者同士(すなわち英語母語話者同士または日本語母語話者同士)の評定値の相関が最も強いことがわかる(相関係数 0.70 - 0.81)。また、母語を共有しない話者間の場合、英語母語話者と非母語話者間の相関(相関係

[1] 評定尺度が小さいと評定値が離散量の性質を帯び、統計解析の際に支障をきたす可能性があるため、Saito et al. (2015)ではスライダーを用いて評定値を 100 段階にしている。本研究では各々の項目の評定値は 10 段階であるが、各話者のデータを 3 分割して、さらに 4 項目の平均値を各話者の評定値とすることにより、一人の話者に対して 10(尺度) x 4(項目) x 3(ファイル)の 120 尺度の評定値を付与することができる。

[2] 音声学的にはイントネーションは「韻律」の下位範疇であるが、音声学的知識から客観的に判断できるアクセントや母音挿入に比べて、直感的に判断されることが予想されたため、主観評価の項目である「流暢さ」に含めた。ただし、作業後の評定者の意見によると、この分類はやや混乱を生じるものだったようで、本研究の不備と言える。しかしながら、イントネーションは 4 つあるうちの 1 項目の、さらに細分化された下位項目であるため、評定結果に大きな影響を与えることは考えにくい。

数 0.62 - 0.80)の方が、非母語話者間(すなわち今回の分析では日本語母語話者-スペイン語母語話者間)の相関(相関係数 0.58 - 0.74)よりも高いという傾向が示された。

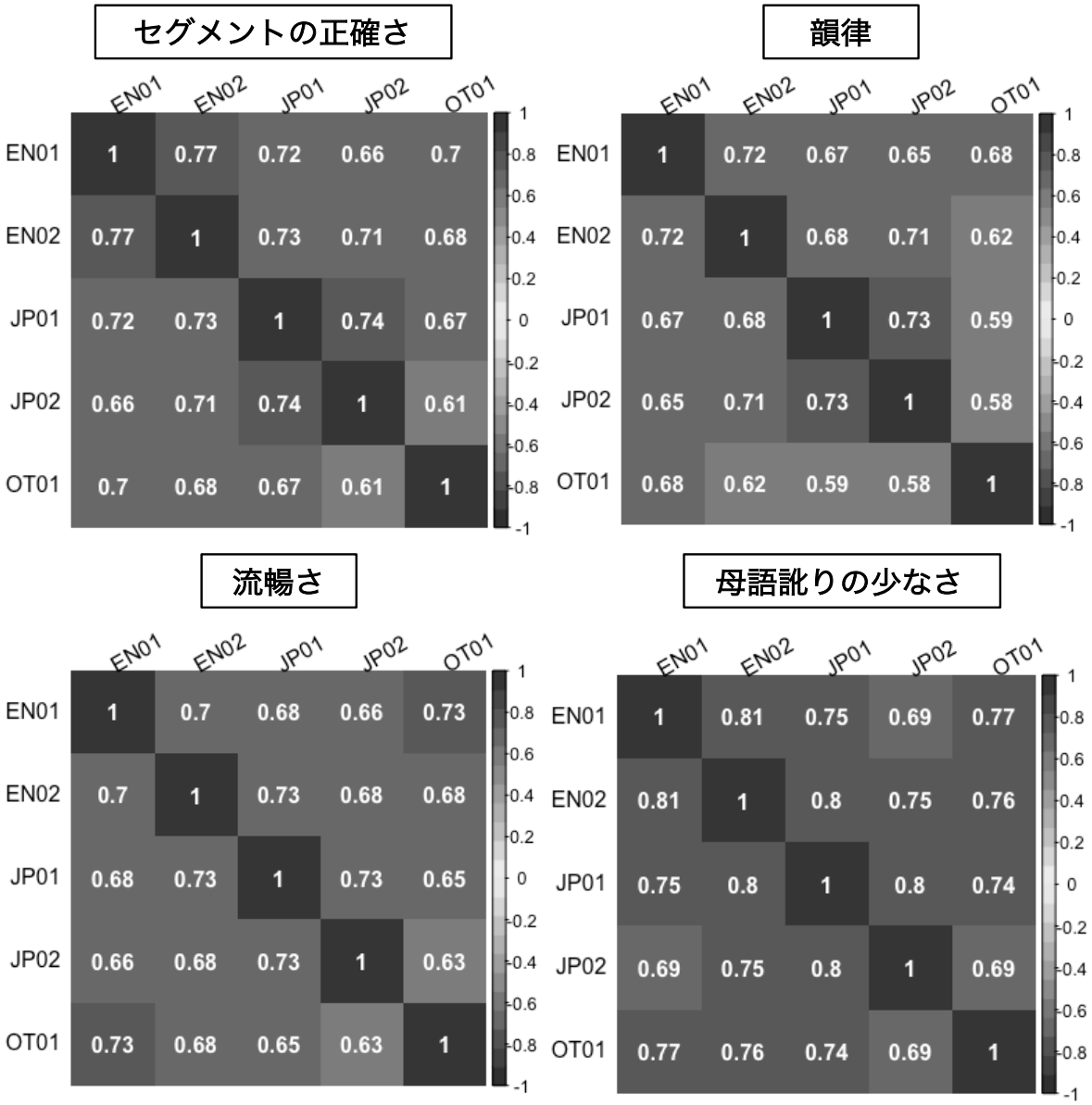


Figure 1 多重相関分析の結果

EN:英語母語話者　　JP:日本語母語話者　　OT:スペイン語母語話者　　01,02:評定者 ID

次に、各項目の評定者間相関を比較してみると、母語訛りに関する評定値の相関係数(0.69 - 0.81)が最も高いことがわかる。母語訛りに関する評定は今回の評定作業において最も直感的に判断される評定項目であることから、音声学を専門とする評定者であっても、音声学的知識を用いた評定よりも直感的な評価の方が一致率が高くなることが示唆される。

また、セグメントに関する評定値一致率(相関係数 0.61 - 0.77)が韻律に関する評定値一致

率(相関係数 0.58 - 0.73)よりも高かったことから、英語のセグメントに関する指標は母語の異なる学習者間においても比較的に共有ており、また、それは英語母語話者のものとも近い一方で、韻律に関する判断は母語の影響を受けやすいという傾向が示された。

5. 今後の展望

今回の分析は既に評定作業を終えた 5 名の評定者のデータを用いて行った比較的に小規模なものであったが、本研究の結果から、L2 英語の評定において英語母語話者間の高い評定値一致率が得られるという傾向は示された。しかしながら、日本語母語話者間の高い一致率は、評定者同士の母語の一致による影響か、または、評定者の母語と評定対象だった L2 英語学習者の母語が一致に起因するものか、どちらの要因が大きいかは今回の分析結果から明らかにすることはできない。今後、他の評定者による評定が終わった時点で再度分析を行い、日本語母語話者以外の非英語母語話者同士(例えばスペイン語母語話者同士)の評定値一致率を日本語母語話者同士の評定値一致率と比較することにより、評定者と異なる母語を持つ L2 英語話者の音声でも、評定者の母語が一致していれば評定値一致率が高くなるかどうかを検証する。

6. 謝辞

本研究は科研費若手研究(B) (No. 17K13513)および基盤研究(B) (No. 15H02729)の助成を受けている。

参考文献

Bolton, K. (2004). World Englishes. In A. Davies & C. Elder (eds), *Handbook of Applied Linguistics*. Oxford: Blackwell, 367-396.

Crystal, D. (2003) *English as a Global language*. Cambridge: CUP.

Saito, K., Trofimovich, P. & Isaacs, T. (2015) “Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels”, *Applied Psycholinguistics* 37(02), 217-240

D5

日本語での中和環境における英語音の知覚 ー *Who* is *Fu*?

青柳 真紀子（獨協大学）Yue Wang (Simon Fraser University)
aoyagi@dokkyo.ac.jp, yuew@sfu.ca

1. はじめに

本研究は，母語の音韻規則が外国語音の知覚におよぼす影響を探るものである．具体的には，日本語話者が英語音/ f /と/ h /を同定する際に，日本語の音韻規則（fu-hu の中和）の影響を受け，/_u/の環境で難しくなることについて取り上げ，その要因を検討する．

外国語音を聞く際，母語における最も近い音素に知覚同化させることが知られている（Best & Tyler, 2007）．簡単に言えば，外国語の二つの音素が母語でも二つの音素に対応していると聞き分けは容易になり，ひいては産出も習得しやすい（Flege, 1996）．英語の/ f /と/ h /は，それぞれ日本語の/ ɸ /と/ h /に知覚同化され，/ f /-/ h /の対立は概して/ ɸ /-/ h /として保たれる（例：ファ vs. ハ）．一見すると，/ f /と/ h /の聞き分けは容易に思われる．ところが，日本語/ hu /は/ h /→[ɸ]となり，/ ɸ /と/ h /ともに[ɸ]として中和し，区別がなくなる（フ vs フ）。これにより，日本語話者にとって *fah-hah*（/ fɑ /-/ hɑ /）の同定は容易である一方，*Fu-who*（/ fu / - / hu /）の同定は困難となり得る．

この/ fu /-/ hu /の中和が絡む困難についての記述は見られるが（竹林, 1997; Kondo *et al.*, 2015 等），知覚面での実証的な研究は多くない．そのうち，Lambacher *et al.* (2001) は日本人大学生の/ f /と/ h /の同定率が u の環境で低いことを示したが，その要因についてはほとんど触れていない．このように，母音環境が影響することは分かるが，そもそも日本語話者が摩擦音/ f /-/ h /そのものの違いが判別できるかどうかはよく分かっていない．

本研究では，まず，/ f /と/ h /を含む語の同定に加えて/ f /と/ h /単独の同定を検証をし，次に，/_u/の環境/ fu /, / hu /から抽出した/ f /-/ h /の同定能力が，元の語全体/ fu /, / hu /，つまり/u/の提示でどのように変化するかを検証する. これにより, 日本語話者が示す/ fu /, /hu/の同定の困難について, 摩擦音の音声処理とそれに対する音韻知識関与の点から考察する．

2. 先行研究

2.1. 無声唇歯摩擦音/ f /と無声両唇摩擦音/ ɸ /

英語等の/ f /は日本語において/ ɸ /で代用されることが多い．現代日本語では/ ɸ /は/ hu /の異音[ɸu]としてウ段でのみで存在するが，外来語では/ f /の代用子音としてすべての母音とともに使われている[1]．このように，/ f /は/ ɸ /で代用され，英語の/ f /-/ h /の対立は概して保たれる（例: ファ vs ハ）．しかし，既述のように/ ɸ /は元来/_u/の環境での/ h /の異音であることから，/ hu /と/ ɸu /はともに[ɸɯ]として中和し，対立を失う（フ vs フ）．唇での摩擦の有無や程度は対立を作らないので, [h]と[ɸ]は相互交換的で音声的には両者の間のどのあたりでも出現する．このように，日本語において調音的にも聴覚的にも区別す

[1] / f /を u の母音つき/ ɸu /で代用する場合もあるが（例: fan > ɸ*u*.an フアン，film > ɸ*u*.i.ru.mu フイルム，felt > ɸ*u*.e.ru.to フエルト），今日，u が二重母音化，さらには脱落して，/ ɸ /は単独で/ f /の代用子音として機能することが多い（例: fan > ɸan ファン，film > ɸi.ru.mu フィルム，felt > ɸe.ru.to フェルト）．

る必要がないことから，これが転移して英語においても/ fu /と/ hu /の同定は難しくなり，これは/ fa/と/ ha /など他の母音の場合の同定と比べても明らかである．

2.2. 無声唇歯摩擦音/ f /と無声喉頭摩擦音/ h /

英語の/ f /は、音響的にはスペクトルピークや傾きを含むスペクトル情報，振幅，CV 振幅比，持続時間，CV 推移などの点から特徴づけられている（Jongman *et al.*, 2000）．しかし，各種摩擦音間の音響的比較および知覚的比較の研究にはほとんど/ h /が含まれない（McMurray *et al.*, 2011 等）．/ h /は後続母音の無声開始部のようなもので特定の狭窄点をもたないことから、摩擦音としてのある一定の音響的特長に欠けるためであろう（Ladefoged, 2005）．このように音響的にも知覚的にも/ f /と/ h /が直接比較されることは多くなく，ましてや/ fu /と/ hu /限定の環境ではほとんど見られない．

日本語話者の英語発音の注意点として/ f /はよく上げられ，両唇[ɸ]ではなく唇歯[f]の摩擦であることが指導される（牧野, 2005 等）．その際/ h /との混同が触れられることもあるが, /_u/の環境/ fu /-/ hu /での違いへの注意喚起や, その知覚についての言及は多くない．

調音面において，Kondo *et al*. (2015) は，日本語話者にとって英語の/ f /-/ h /は/ fu /-/ hu /の音素配列に特化して困難であるとしている（Table 2(b)）．実際，96 人による読上げ発話において/ f /と/ h /の混同がよく見られることが示されているが，混同の母音環境については特に触れられていない. Nogita (2010) も日本人英語学習者 4 人の英語読上げ実験で *food-who'd* における/ f /と/ h /の混同を示している．

本題の知覚面では，Lambacher *et al.* (2001) が，日本人大学生 100 余名に対する英語子音聞き取り実験により，/ f /と/ h /は，u の環境/_u/，/u_u/，/u_/で同定正答率と識別感受性（*d'* スコア）が低いことを示しているが，その要因については分析されていない．

3. 実験

3.1. 目的

日本語での中和環境（/ ɸu /-/ hu / = フ vs フ）に対応する英語音の対立/ fu /-/ hu /の知覚について，日本語話者と英語話者を対象に調査をする．まず，/ f /と/ h /を含む語（実験 1），および語から抽出した摩擦音部分のみ（実験 2）について語彙，子音の同定実験を行う．次に，語から抽出した母音部分のみについて，先行子音の検知実験を行う（実験 3）．

実験 1 で/_u/の環境の影響を確認した後，実験 2 で摩擦音単独での同定，さらに，実験 3 で後続母音部に残る情報による先行子音の検知について検証する．焦点は，/ fu /と/ hu /から抽出した/ f /と/ h /の摩擦音の同定が，元の語全体/ fu /と/ hu /の提示，つまり/_u/の環境の提示によってどのように変化するかを見ることである．これにより，母音環境を限定した中和という音韻規則が，子音そのものの知覚に与える影響を考察する．

3.2. 方法

＜使用した語＞ 実験には，表 1 に示すように/ f /と/ h /から始まり，/_i /，/_u /，/_a /，および問題の母音/_u /との境界を示すための/_oʊ /を含めた 4 母音が後続する 1 音節対語，さらに，渡り音をはさんで/ u /が続く（/_ ju /）1 音節対語の計 5 対 10 語を使用した．また，*see – she, sip – ship, said – shed, sue – shoe, sock – shock* をダミーとして使用した．

表 1. 使用した語[2]

	/ i /	/ u /	/ ɑ /	/ oʊ /	/ ju /
/ f /	*fee*	*Fu*	*fah*	*foe*	*few*
/ h /	*he*	*Who*	*hah*	*hoe*	*Hugh*

＜録音＞ 成人北米英語母語話者 4 人（男 2，女 2）が実験語を単独で読み上げたものを録音した（マイク Shure KSM，サンプリング周波数 44.1kHz，量子化 16bit）.

＜刺激音＞ 以下の 3 種類の刺激音を作成した.

実験 1: 語全体　表 1 の 10 語．語全体を 65dB に標準化した.

(1) *fee-he,　Fu-who,　fah-hah,　foe-hoe,　few-Hugh*

実験 2: 摩擦音部　(a) 上記 *fee-he, fah-hah, Fu-who* から摩擦音部[f]と[h]を抽出したもの，および (b) [f]と[h]を単独で発したもの，計 8 刺激．(a)は波形上で摩擦開始部分から後続母音の最初の有声周期開始部分までとした．摩擦部全体を 45dB に標準化した.

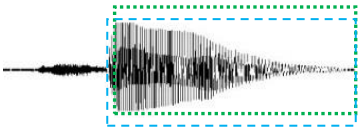
(2a) *f*(ee),　*h*(e),　*f*(ah),　*h*(ah),　*f*(u),　*wh*(o)
（カッコ内は分離された母音）

(2b) *f*(iso),　*h*(iso)
（iso: 単独での発話）

実験 3：後続母音部　実験語のうち *fee-he, fah-hah, Fu-who* から上記(2)の摩擦音部分を除いた母音部分で波形上，(a) 母音の 1 つ目の周期への立ち上がりから母音終了点までと，(b) 3 つ目の周期への立ち上がりから母音終了点までを抽出したもの、計 12 刺激．(a)の方が，先行子音（f または h）の同時調音的情報をより多く含んでいると考えられる.

(3a) (f)ee_1stV,　(f)ah_1stV,　(f)ah_1stV,　(h)ee_1stV,　(h)ah_1stV,　(h)ah_1stV

(3b) (f)ee_3rdV,　(f)ah_3rdV,　(f)ah_3rdV,　(h)ee_3rdV,　(h)ah_3rdV,　(h)ah_3rdV
（カッコ内は分離された子音）



＜手順＞　作成された刺激は以下のとおり実験ソフト *Paradigm* 上に準備された.

実験 1 語全体：計 160 刺激（[10 語 x 4 話者] x 2 回＋ダミー80 刺激 ＝160）

実験 2 摩擦部のみ：計 96 刺激（[2 子音 x 4 母音環境 x 4 話者] x 3 回 ＝96）

実験 3 母音部のみ：計 96 刺激（[3 母音 x 2 子音環境 x gating2 段階 x 4 話者] x 2 回= 96）

刺激は被験者毎にランダムに提示され，直後に選択肢（例: fee　he）が画面に表示される．被験者はヘッドセットで刺激を聞き，画面上で強制二者択一の同定判断をする.

被験者は日本語母語話者大学生 14 人（男 6，女 8）と北米英語母語話者大学生 14 人（男 2，女 12）で，手順説明の際に実験語の意味と音，特に/ f /と/ h /の対語群（*who* は綴りは‘wh’でも/ h /で始まる語）であることを確認した．練習と休憩を入れて合計 45-50 分であった.

[2] *Fu* は中国人名（*Mr. Fu*），*fah* は音符名（ドレミ「ファ」），*hah* は間投詞，*Hugh* は人名.

4. 結果と考察

4.1. 実験１（f と h を含む語全体の同定）

まず, 英語母語話者（EngNS）の f と h の対語の正答率は母音環境にかかわらずほぼ 100%であった（*hah* は 99%）．これに対し，日本語話者（Jpn）は *Fu/who* と *few/Hugh* の正答率が低い（図 1）．*Fu/who* は予想通りだが，*few/Hugh* は日本語でも対立があるので（フュー/ヒュー）予想外であった．語彙親密度の影響だと思われる反応が見られ，*Fu/who* では聞き慣れない *Fu* に対し頻出の *who* (/ h /)の回答率と正答率が高く，逆に *few/Hugh* では不慣れな *Hugh* に対し頻出の *few* (/ f /)が高い．対語間の反応の偏りが関与することから[3]，信号検出論に基く識別感度（*d'*）を見ると（図 2），/_u/と/_ju/で同定が困難であることが分かる．Jpn 上級者（TOEIC850 以上）の *few/Hugh* は比較的よいことからも，*Fu/who* の困難さが確認される．

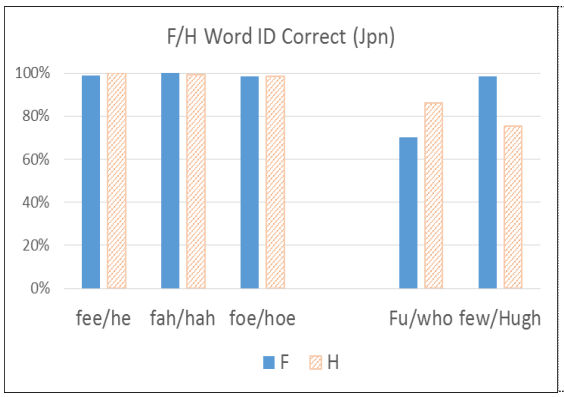


図 1. f/h 語彙同定の正答率 (Jpn)

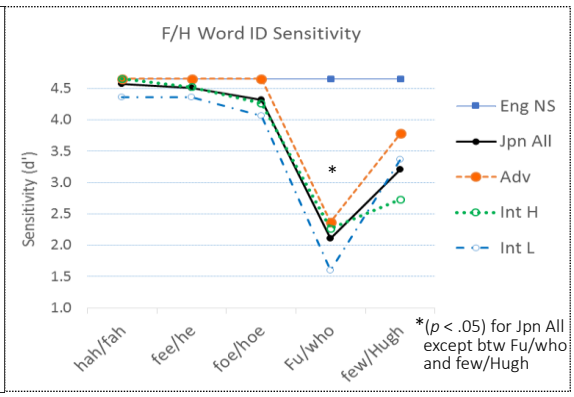


図 2. f/h 語彙同定の識別感度 (EngNS, Jpn)

4.2. 実験２（抽出された f と h の同定）

摩擦音部分のみでも EngNS の正答率と識別感度は母音環境に関係なくほぼ完璧である．Jpn は，*Fu/who* から抽出された *f*(u)-*wh*(o)，および単独発話された *f*(iso)-*h*(iso)で感度が低下するが，全般として摩擦音のみでもかなり子音を同定できることがわかる（図 3, 4）．本研究の焦点である語全体と抽出子音単体での感度の差については，別に 4.4 節に示す．

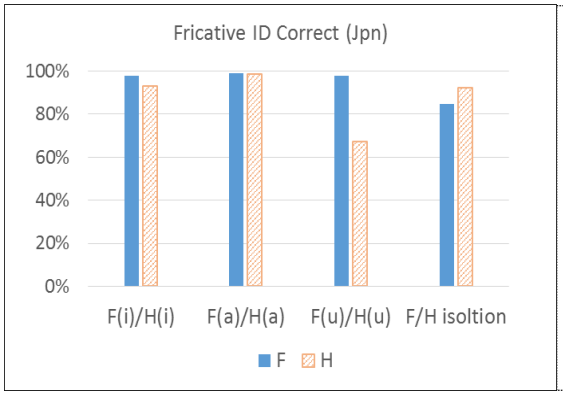


図 3. 抽出された f/h の同定正答率 (Jpn)

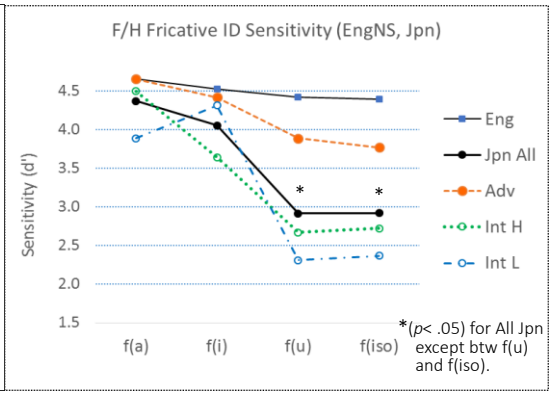


図 4. 抽出された f/h の識別感度 (EngNS, Jpn)

3 仮に常に/f/と回答すると、f の正答率は 100%になるが h は 0%となる．この f の 100%は f/h の完璧な判別力を意味しない．一方，f の正答率が 100%で h も 100%であれば，f の 100%の意味合いは先の例とは全く異なる．よって，f と h 単体での正答率ではなく，両者の正答と誤答の兼合いが識別感度には重要となる．

4.3. 実験３（抽出された母音部によるｆとｈの検知）

母音部から先行子音 f/h 検知する課題の正答率は概して低くなるが，EngNS では全母音で，Jpn でも(f)*u* の場合を除いて偶発率以上となる．識別感度を見ると，EngNS は母音にかかわらず，母音が進んで先行子音の情報が減少するにつれ子音検知が難しくなる自然なパタン見られるが（図 5），Jpn は(f/h)*u* ではこれが少なく，(f/h)*e* を除いて入渡り部の音響情報への感度が鈍い。特にここでも/_u/の環境 (f/h)*u* での感度の悪さが見られる（図 6）．

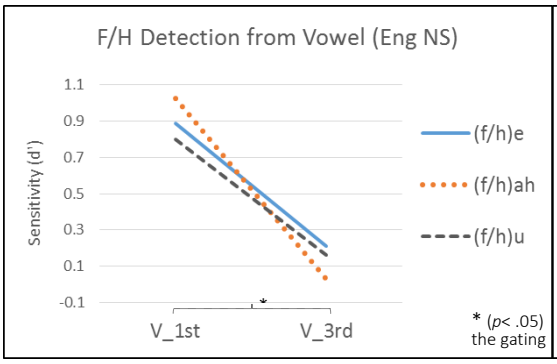


図 5. 母音から先行 f/h を検知 (EngNS)

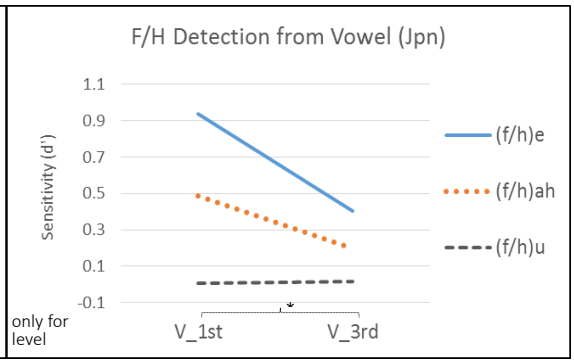


図 6. 母音から先行 f/h を検知 (Jpn)

4.4. 子音のみ（C）と子音-母音（CV）の知覚

通常，子音の同定は子音のみ（C_1）よりも後続母音を伴った（C_1V）方が容易である（$C_1 < C_1V$）（C_1V と C_1 の C_1 は抽出前後の同じ子音）．例えば，[f]単独よりも[fa]の方が推移部のキューなど音響的情報が多く，また音節語として成立し，経験と予測からくる処理能力の向上も期待される（McMurray *et al.*, 2015）．ところが，Jpn は，/_u/の環境では子音のみの方が識別感度が高くなる（C > CV）（表 2 下段）．つまり，*fah/hah* や *fee/he* から抽出された[f], [h]の識別感度は，元の語全体が提示されると上昇する一方で，*Fu/who* から抽出された[f], [h]では，単体ではできていた識別が，それが元々含まれていた抽出元の語全体が提示されるとできなくなってしまうと言える．一部のみでできていた識別が，それを含む全体の提示によってできなくなるということは，一般的な予測に反する．これは，音声的に[f]と[h]がある程度識別できていても，/_u/の提示により/ f /と/ h /は異音になるこ

表 2. “抽出された C の同定”と“抽出元の語全体の同定”の感度比較（Jpn）

	C f / h		CV fah/hah
Jpn All	4.37	↗	4.57
Advanced	4.65	↗	4.65
Intermed H	4.50	↗	4.65
Intermed L	3.89	↗	4.36

	C f / h		CV fee /he
Jpn All	4.05	↗	4.51
Advanced	4.42	↗	4.65
Intermed H	3.64	↗	4.52
Intermed L	4.31	↗	4.36

	C f / h		CV Fu/who
Jpn All	2.91	↘	2.10
Advanced	3.89	↘	2.37
Intermed H	2.67	↘	2.26
Intermed L	2.31	↘	1.60

↗：C より CV で感度上昇
↘：C より CV で感度低下

とから，異音識別の鈍化が引き起こされている可能性が考えられる．また，これは Boomershine *et al.* (2008) の示した音素と異音の知覚感度の違いと類似するものである．

5. 結論と今後

これまで見てきたように，語全体でも，摩擦音部のみでも，母音部のみでも，日本人被験者は他の母音環境と比べて/_u/の環境で/ f /-/ h /の識別感度が低い（それぞれ 4.1, 4.2, 4.3）．これに加え，/_u/の環境で，同じ信号 C_1 の同定感度がその抽出元全体 C_1V で鈍化すること（$C_1 > C_1V$，つまり *f*(u)-*h*(u) > *Fu-who*）が見られたことは（4.4），/ f /-/ h /ついて，より基礎的な音声処理が，より高次の音韻規則の知識によって阻害されていることを示すと考えられる．また，通常 C_1V は C_1 の知覚を容易にするが，逆に阻害する一例でもある．

別の見方として，日本語話者は摩擦音/ f /-/ h /のみの識別時には音声として聞いていない可能性がある．Werker *et al.* (1985) が，英語話者がセイリッシュ語のある種の子音を言語音声としては同定できないが非言語音声としては同定できることから別モードでの知覚を示唆していることに類似するが，本研究のデザインではそれを検証するに至らない。

また，/_u/の環境の提示が/ f /-/ h /の異音化と識別鈍化を引き起こすという仮説を検証するには，弁別精度を見る実験，子音部分に母音情報を加算していく実験，また，/ _u /環境の擬似提示（音響ではなく「期待」の提示）によるプライム効果を確認する実験等，別の手法が必要となろう．現在，/f_/-/h_/の音響差異の母音環境による変化を分析中だが，この結果とそれに基づいた知覚実験も音響と知覚の関係を探る基礎項目として必須である．

謝辞：本研究の一部は Simon Fraser University, Language and Brain Lab と Department of Linguistics の協力と助成を受けて行われ，また，河合裕美先生にも実験実施にご協力いただきました。ここに感謝します．

参考文献

竹林滋 (1997)『英語音声学』東京：研究社.

牧野武彦 (2005)『日本人のための英語音声学レッスン』東京：大修館.

Best, C., & Tyler, M. (2007) “Nonnative and second-language speech perception,” In O.-S. Bohn & M. Munro (eds.) *Language Experience in Second Language Speech Learning.* Amsterdam: John Benjamins.

Boomershine, A., Hall, K., Hume, E. & Johnson, K. (2008) “The impact of allophony vs contrast on speech perception,” In B. Avery, E. Dresher & J. Rice (eds) *Phonological Contrast*. N.Y.: Mouton de Gruyter.

Flege, J. E. (1996) “Second language speech learning: Theory, findings, and problems,” In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-linguistic Research*. Timonium: York Press.

Jongman, A., Wayland, R. & Wong, Serena. (2000) “Acoustic characteristics of English fricatives,” *J. of Acoustical Society of America* 108 (3), Pt. 1, 1252-1263.

Kondo, M., Tsubaki, H. & Sagisaka, Y. (2015) “Segmental variation of Japanese speakers’ English: Analysis of “the North Wind and the Sun” in AESOP Corpus,” *J. of Phonetic Society of Japan* 19 (1), 3-17.

Lambacher, S., Martens, W., Nelson, B. & Berman, J. (2001) “Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias,” *Acoustical Science and Technology* 22 (5), 334-343.

McMurray, B. & Jongman, A. (2011) “What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations,” *Psychological Review* 118 (2), 219-246.

McMurray, B. & Jongman, A. (2015) “What comes after /f/? Prediction in speech derives from Data-Explanatory Processes,” *Psychological Science OnlineFirst*, published on Nov. 18, 2015.

Nogita, A. (2010) “Do Japanese ESL learners’ pronunciation errors come from inability to articulate or misconceptions about target sounds?” *Working Papers of the Ling. Circle of U. of Victoria* 20 (1), 82-116.

Werker, J. & Logan, J. (1985) “Cross-language evidence for three factors in speech perception,” *Perception & Psychophysics* 37 (1), 35-44.

D6

韓国語母語話者による英語の母音の知覚判断
-後続子音の影響について-

韓 喜善（大阪大学）, 野澤 健（立命館大学）
kenkyuhhs@gmail.com, t-nozawa@ec.ritsumei.ac.jp

1. 本研究の背景と目的

これまで、英語の母音に関する知覚実験の多くは、子音の影響を最小限にするために、声門摩擦音と閉鎖音の間に母音を入れた音環境を中心に検討されてきた。また、非母語話者を対象とした場合、母音に後続する子音の影響について調査した論考は少ない。本研究では、韓国語母語話者による英語の/ i, ɪ, ɛ, æ, ɑ, ʌ /の 6 つの母音の知覚判断について調査した。その際、閉鎖音/t/、鼻音/n/、流音/l/を後続子音として配置し、後続子音の調音様式による母音の知覚判断の違いについて検討した。

2. 先行研究

2.1. 韓国語と英語の短母音

母音の数や種類に関しては、研究者によって見解が異なる場合もあるが、現代の韓国語の短母音は/i, e, a, o, u, ʌ, ɯ/の 7 つという見解（어 1996、Kang 2013、Hwang 2015 等）が主流である。米英語の短母音については、Ladefoged (2003, 2006)では、/i, e, ɪ, ɛ, æ, ɑ, ʌ, ʊ, u,o, ə, ɚ/の 12 個という見解が示されている。両言語の母音は、母音の音素の数が異なるだけでなく、互いの母音の舌の前後や舌の高さの配置が異なっている（전 2005）。

2.2. 韓国語母語話者による英語の母音の生成と知覚

英語の/i, ɪ , æ, ɛ/の母音はすべて前舌母音ではあるが、前舌母音（F2）としての度合いに差があり、口の開き（F1）の度合いも異なる（Ingram and Park 1997、Ladefoged 2006 等）。そのため、これらの子音を聞き分けるためには、それぞれの F1 と F2 の周波数に注目する必要がある。韓国語母語話者の英語の母音の知覚判断については、/i/と/ɪ/との間、/æ/と/ɛ/との間の弁別が困難であることが報告されており（Ingram and Park 1997、Flege et al. 1997、Frieda and Nozawa 2007 など）、/i/と/ɪ/を韓国語の/i/として/æ/と/ɛ/を韓国語の/e/に知覚しているという（Frieda and Nozawa 2007）。生成においても、Ingram and Park (1997)によると、韓国語母語話者の/i/と/ɪ/の F1 と F2 は同等であり、/æ/と/ɛ/の F1 と F2 も同等であると報告している。これらの母音の判断には、母音長に注目すれば、母音の判断率が高くなることが日本語母語話者の例から知られているが（Ingram and Park 1997、Frieda and Nozawa 2007）、韓国語母語話者は学習歴や現地滞在歴に関わらず、このようなストラテジーを採用しないため、英語の母音の知覚と生成の問題は継続する傾向があるという。

一方、後舌の/ɑ/と中舌寄りの/ʌ/については、韓国語母語話者がこれらの母音の弁別がほぼ可能であることも報告されている（Frieda and Nozawa 2007）。

このように、韓国語母語話者による英語の/i, ɪ , æ, ɛ/の音声の難易度に関しては、先行研究間で一貫した結果を得ているが、いずれも後続母音が閉鎖音で終わる環境での報告に限られている。

3. 本研究の課題

本研究では、先行研究で取り上げた英語の 6 つの母音（/i, ɪ , æ, ɛ, ɑ, ʌ/）について、後続する子音の影響を調査する。後続子音として閉鎖音（/t/）、鼻音（/n/）、流音（/l/）を選び、それぞれの環境を有する英語の母音の弁別について比較した。先行研究で取り上げられていた語末閉鎖音の他に鼻音と流音を調査項目として加えたのは、母音に後続する鼻音のフォルマントが母音の F1 付近に現れることにより、母音の高さの知覚が影響されるものと予想されるためである。特に、/æ/は鼻音の前では上方（F1 の周波数の低下）、前方（F2 の周波数の上昇）に移動することが知られているが（Labov 2001）、このことが韓国語母語話者の英語の知覚に影響することも予想される。また/l/が後続すると F2 が継続して降下するため、他の子音環境とは異なる響きがすることが予想される。これらの仮説について調査を行い、その検証を行う。

4. 実験の手順

4.1. テスト語

米英語の母語話者 4 名（20 代の女性、大学生）による音声を使用した。4 名の話者の出身地は、ニューヨーク州 2 名、カリフォルニア州 1 名、ウィスコンシン州 1 名である。オーバーン大学（Auburn University、米国アラバマ州）の実験室内のコンピュータの HD に直接録音した。/hVt/、/pVt/、/pVn/、/pVl/の環境で発した 24 語を刺激音の素材として使用した。テスト語の中には、人名（Pete、Pell）や無意味語（het、pon、pul）も含まれる。

表 1　テスト語のリスト

	i	ɪ	ɛ	æ	ɑ	ʌ
hVt	heat	hit	het	hat	hot	hut
pVt	Pete	pit	pet	pat	pot	putt
pVn	peen	pin	pen	pan	pon	pun
pVl	peel	pill	Pell	pal	pol	pul

4.2. 音響分析

収集した音声は、母音長、母音の中間時点での F1 と F2 を測定した[1)]。長さに関しては、/i/は/ɪ/に対して 1.1〜1.3 倍長く、/æ/は/ɛ/に対して 1.1〜1.7 倍長く、/ɑ/は/ʌ/に対して 1.1〜1.6 倍長かった。特に/pVn/の環境が母音種による長さの割合の差が他より大きかった。

[1)] /pV1/の環境に関しては母音と/l/との分節ができないため、長さに関しては母音と/1/の区間を測定し、フォルマントに関しては母音と/l/区間の 1/4、2/4、3/4 の時点についてそれぞれ F1 と F2 の測定を行った。

フォルマントに関しては、/hVt/と/pVt/とでは、それぞれの母音の F1 と F2 がほぼ同様の周波数を示し、語頭子音の調音位置や調音様式による影響はなかった。

/pVn/では、/hVt/と/pVt/に比べて、/i, ɪ, ɛ, æ/は F2 が高く、/ɑ/と/ʌ/は F2 が低かった。特に、/æ/の F2（2407Hz）は/hVt/（1789 Hz）と/pVt/（1897 Hz）の環境より著しく高く、Labov (2001)の結果と一致する。F1 に関しては、/i, ɪ, ɛ, ʌ/は/hVt/と/pVt/の環境の F1 とほぼ同じかそれに近い周波数を示しているが、/æ, ɑ/では/pVn/の環境においては F1 の周波数が低くなっている。/pVn/は、/pVt/環境と調音場所が一致するため、調音様式（口音か鼻音か）による影響と推測できるが、今後さらに詳しい検討が必要である。

/pVl/の母音と/l/区間の 4/1 時点については、すべての母音の F2 の周波数が低いことから、後寄りに発音されていることがわかる。これは英語の語末に位置する/l/の音声がいわゆる“dark l（軟口蓋化した有声歯茎側面接近音：[lˠ]、軟口蓋化有声歯茎側面接近音：[ɫ]）”として発音されることによって、後舌が軟口蓋に接近するという調音が先行する母音の調音段階から始まっているものと解釈できる。舌の高さについては、母音によって傾向が異なり、/i, ɪ/は F1 の周波数が多少高くなり広母音化している。それに対して、/æ, ɑ, ʌ/は F1 の周波数が低くなり、狭母音化している。/ɛ/は/hVt/、/pVt/の場合とほとんど変わらない。このように、/l/は先行する母音を後舌化すると同時に、狭母音と広母音の差を不明瞭（結果として中舌化）なものにする働きをしていることがわかる。この傾向は子音部への移行が進むほどより顕著になっていく。

4.3. 実験方法

実験は、/hVt/、/pVt/、/pVn/、/pVl/のそれぞれ音環境ごとに行った。実験参加者に、音声刺激を聞き、聞いた母音を選択肢の中から選択する方式で回答させた。選択肢は「4.1.」で説明した通り（表 1 を参照）、/i, ɪ, ɛ, æ, ɑ, ʌ/の 6 つのいずれかの母音を含む有意味ないし無意味語である。提示した語は音声記号ではなく、英語の綴りで表記したため、音と綴りとの関係を予め説明した。例えば、het は pen と同じ母音であること、putt は put ではなく but と同じ母音を含む語であることを説明した。

刺激音は 2 回試行し、音環境ごとに 48 個（6 母音×4 話者×2 回=48 個）の刺激音に対する評価となる。刺激音はランダムに配列し、子音環境の提示順による影響を考慮し、参加者により、子音環境の提示順番を変えた。4 つの子音環境で計 192 個の刺激音の判断を 30 程度所用した。途中で休憩を 10 分ほど入れた。

4.4. 実験参加者

実験参加者は韓国母語話者 22 名（18-35 歳、男性:6 名、女性：16 名）である。すべての学習者は韓国語を母語とし、近畿圏の大学に在籍している学部生と大学院生である。日本滞在歴は 1 年 3 ヶ月〜15 年 3 ヶ月である。英語能力に関しては 1〜9 段階（1：very poor、9：very good）の中から選ぶように指示した。

5. 結果と分析

6 母音の各子音環境での平均正答率を表 2〜5 に提示する。4 つの子音環境と 6 母音を被験者内要因にした反復測定の分散分析の結果、子音環境の主効果（$F(3, 63)= 41.782$, $p<.001$）、母音の主効果（$F(5, 105)=35.469$, $p<.001$）、子音環境と母音の交互作用（$F(15, 315=19.140$, $p<.001$)といずれも有意であった。多重比較の結果、/i/を除くすべての母音が子音環境の影響を受けていることがわかった。/ɪ/では、/pVn/は他の音環境より有意に正答率が低く（$p<.001$）、/hVt/は/pVt/より正答率が低い（$p<.01$）。/ɛ/では、/hVt/は他の音環境より有意に正答率が低い（$p<.01$）。/æ/では、/hVt/は他の音環境より有意に正答率が高く（$p<.01$）、/pVl/は/pVt/より有意に低い（$p<.01$）。/ɑ/では、/pVn/と/pVl/は/hVt/と/pVt/に比べて有意に正答率が低く（$p<.001$）、/pVn/より/pVl/のほうが正答率が低い（$p<.01$）。/ʌ/では、/pVl/は他の音環境に比べて有意に正答率が低かった（$p<.001$）。以下、/i/と/ɪ/、/ɛ/と/æ/、/ɑ/と/ʌ/の比較を中心に論を進めていく。

5.1. /i/と/ɪ/

/i/と/ɪ/の音韻の混乱が見られ、先行研究と一致する結果を得た（Ingram and Park 1997、Flege et al. 1997、Frieda and Nozawa 2007 など）。全体的に/i/を含む語の heat（55%）、Pete（47%）、peen（42%）、peel（47%）に対する正答率より/ɪ/を含む語の hit（67%）、pit（84%）、pin（39%）、pill（75%）に対する正答率が高い。実験後に、実験参加者に heat と hit、Pete と pit、peen と pin、peel と pill の区別をどのようにしたかについてインタビューをしたところ、「母音の長さ」という回答が多かったが、実際には母音長に差があっても（4.2.参照）、母音長の影響とは言えない結果となっている。特に、/pVn/では母音長の相対的な差が他の環境より顕著であることを「4.2.」で報告した。このように、/i/を/ɪ/に間違えやすいという結果に偏ったのには、綴り字の影響を考慮する必要がある。つまり、例えば heat か hit か判断に迷った際に綴りがより簡単な hit を優先的に選んだ可能性がある。/pVn/の環境においては、/ɪ/を/ɛ/と間違える場合があった（24%）。/ɪ/と/ɛ/の F1 の差は、/hVt/と/pVt/に比べて/pVn/の環境で小さい。

5.2. /ɛ/と/æ/

/æ/と/ɛ/についても、これら 2 つの音韻を間違いやすいという結果となり、先行研究と一致した（Ingram and Park 1997、Flege et al. 1997、Frieda and Nozawa 2007 など）。音環境によって以下の 2 つの傾向が見られた。/ɛ/の正答率は、/hVt/（48%）が/pVt/（63%）、/pVn/（69%）、/pVl/（77%）よりむしろ低い。しかし、/æ/の正答率はその反対で/hVt/（69%）で最も高い。これは、有意味語（het）か無意味語（hat）かによる影響を考慮する必要があり、/ɛ/か/æ/かの判断に迷った時に有意味語を優先して選んだ可能性がある。したがって、以降では音環境の影響については/pVt/、/pVn/、/pVl/の間の比較に絞って検討する。

/æ/の正答率は、/pVt/（47%）より/pVn/（39%）と/pVl/で低い（32%）。/pVn/で/æ/を/ɛ/と間違えやすかった理由は、/æ/は鼻音の前では上方（F1 の周波数の低下）と前方（F2 の周波数の上昇）に移動し、/ɛ/に近づいたためと解釈できる。一方、/pVl/でも/æ/は/ɛ/に回答が集中しているが、ここでは/ɛ/の F2 が/æ/に近づいているためと解釈できる。しかし、/ɛ/の正答率に関してはその正答率が高く（78%）、不明な点が残る。

5.3. /ɑ/と/ʌ/

音環境によって 2 つの傾向に分かれた。まず、/hVt/、/pVt/の閉鎖子音の環境では、72%以上の割合で/ɑ/と/ʌ/の弁別ができている。これは、先行研究と一致する結果である（Frieda and Nozawa 2007 など）。一方、/pVn/では/ɑ/（無意味語 pon）を/ʌ/（pun：45%）と判断する場合がある。/pVn/では、/ɑ/と/ʌ/がかなり接近していた。この現象は/pVl/の環境においても見られた。また、/pVl/では/ɑ/（poll）を/æ/（pal：49%）か/ʌ/（無意味語 pul：23%）に判断するという結果であった。これは、この環境において/ɑ/と/æ/は舌の高さ（F1）で類似し、/ɑ/と/ʌ/は舌の前後（F2）の位置が近い。その結果、pul（/ʌ/）に関しては/ɑ/（47%）と判断しやすかったと見られる。

表 2　/hVt/の知覚判断の結果(%)

		選択肢					
		/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
刺激音	heat /i/	**55**	**44**	2	0	0	0
	hit /ɪ/	**28**	**67**	2	2	0	1
	het /ɛ/	1	7	**48**	**39**	1	5
	hat /æ/	2	1	**27**	**69**	1	1
	hot /ɑ/	1	0	1	9	**77**	13
	hut /ʌ/	0	0	3	5	10	**82**

表 3　/pVt/の知覚判断の結果(%)

		選択肢					
		/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
刺激音	Pete /i/	**47**	**52**	1	0	0	1
	pit /ɪ/	14	**84**	1	1	0	0
	pet /ɛ/	11	6	**63**	14	1	6
	pat /æ/	2	2	**48**	**47**	1	1
	pot /ɑ/	0	0	0	13	**72**	15
	putt /ʌ/	0	1	2	1	10	**86**

表 4　/pVn/の知覚判断の結果(%)

		選択肢					
		/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
刺激音	peen /i/	**42**	**57**	1	0	1	0
	pin /ɪ/	**21**	**39**	**24**	4	0	0
	pen /ɛ/	1	14	**69**	15	1	1
	pan /æ/	2	2	**57**	**39**	0	0
	pon /ɑ/	0	1	0	10	**44**	**45**
	pun /ʌ/	0	0	1	5	15	**66**

表 5　/pVl/の知覚判断の結果(%)

		選択肢					
		/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
刺激音	peel /i/	**47**	**51**	1	0	0	1
	pill /ɪ/	16	**75**	5	1	1	2
	Pell /ɛ/	3	8	**77**	8	1	3
	pal /æ/	2	0	**63**	**32**	2	1
	pol /ɑ/	0	0	1	**49**	**27**	**23**
	pul /ʌ/	0	1	2	10	**47**	**41**

6.　結論

今回の実験結果は、韓国語母語話者は/i/と/ɪ/、/æ/と/ɛ/の判断が困難であるという点で先行研究と一致する。しかし、/ɑ/と/ʌ/に関してはこれら 2 つの音韻の弁別が可能である（Frieda and Nozawa 2007）ということが知られてきたが、鼻音が後続する環境や流音が後続する環境ではこれらの音素の区別は容易ではないことが明らかになり、音環境によって音素の区別の難易度は変化することがわかった。また、本研究に参加した実験参加者は全員第 2 言語として日本語を習得しており、その大半が母音の長短が習得できていることが期待されたが、実際には英語の知覚判断においては母音の長さに注目していないことがわかった。

参考文献

Ingramm, J. CL and S. Park (1997) “Cross-language vowel perception and production by Japanese and Korean learners of English”, *Journal of Phonetics*25:3, pp. 343-370.

Kang, Y. (2013) “A Corpus-Based Study of Positional Variation in Seoul Korean Vowels,” *Japanese/Korean Linguistics* 23.

Ladefoged, P. (2006) *A Course in Phonetics*, Wadsworth.

Hwang, H. (2015) “The role of perceived similarity and contrast: English loanwords into Korean and Japanese,” *Proceeding of International Conference on Phonetics and Phonology 2015*.

Frieda, E. and T. Nozawa (2007) “You are what you eat phonetically: The effect of linguistic experience on the perception of foreign vowels,” *Language Experience in Second Language Speech Learning in honor of James Emil Flege*, pp. 189-200, John Benjamins.

Labov, W. (2001) *Principles of Linguistic Change*, Blackwell.

Ladefoged, P. (2003)「アメリカ英語」『国際音声記号ガイドブッグ』, pp. 58-62.

이호영(1996) 국어음성학, 태학사.

전상범(2005) 전면 개정판 영어음성학개론, 을유문화사.

本研究は科研費基盤研究 C（16K02650、研究代表者：野澤健）の助成を受けたものである。

P01

中国語を母語とする日本語学習者による感情音声の知覚

李　歆玥（神戸大学大学院）・羅　米良（大連外国語大学）・林　良子（神戸大学）
lixinyue12200@163.com・luomiliang@gmail.com・rhayashi@kobe-u.ac.jp

1. はじめに

音声も表情と同様に、喜び、怒り、悲しみなどの感情の知覚に文化を超えた共通性があるということがたびたび指摘されている(Tickle 2000 他)。一方で、感情の知覚は母語話者と非母語話者の間で異なる傾向があることも報告されている。感情音声の表出には、声の高さ・長さ・強さ・声質などの複合的な音響特徴が関与していることが知られており、各言語によってその音響特徴も異なる(Johnstone 2001)。Campbell & Erickson (2004)では様々な発話音声に対し、自由に評価してラベリングするという実験を行ったところ、異なる母語話者間では同一音声に異なったラベリングを行うことが観察され、このような知覚の相違が感情音声知覚の誤解を引き起こす可能性があると述べている。日本語教育においては、感情音声の聞き取り教材（西端 1996）がわずかながら開発されているものの、感情は普遍的なものであるため指導する必要がないという認識が通用しており、日本語教育の現場において感情の表出、理解が扱われることがほとんどないとされる (中林 2011)。そのため、日本語学習者による感情音声の産出、知覚についてはこれまでにあまり多くの研究が行われてこなかった。

本研究では、このような背景をもとに、中国語を母語とする日本語学習者（以下 JFL）は日本語音声に含まれる感情を一体どの程度知覚できるのか、日本語学習歴により感情音声の知覚が変化するか、また、学習者に混同されやすい感情パタンは何かについて明らかにすることを目的とする。

2. 先行研究

2.1 日本語母語話者による感情音声の産出・知覚

重野（2001）では「東京」「河原崎さん」「11 時半」「さようなら」「そうですか」に含まれる「幸福」「驚き」「怒り」「嫌悪」「恐れ」「悲しみ」という 6 つの感情発話の産出について分析した。その結果、基本周波数とエネルギーの標準偏差が、「怒り」では他の感情より高く、「嫌悪」では最も小さく、大きな差異が認められた。田川（2001）では、第 1 拍目の後に下降が来る「見ないの」という音声の「平静の問いかけ」と「非難の問いか

け」について分析を行った。「非難の問いかけ」の冒頭部分から上昇部分までのピッチ曲線は「平静の問いかけ」より急な傾斜が見られ、「非難の問いかけ」の全体時間長と各音節の時間長は「平静の問いかけ」より長いと述べている。２つの研究のみを挙げたが、これだけでも日本語の感情音声産出には、発話のピッチパターンや持続時間、声の大きさが大きく関与していることが分かる。

2.2 非日本語母語話者による感情音声の産出・知覚

エリクソン・昇地（2006）は、日本語話者が「喜び」「悲しみ」「怒り」「驚き」「疑い」の5種類の感情を込めて発話した「バナナ」を日本人母語話者、アメリカ英語母語話者、韓国語母語話者に聞いてもらい、どの感情に知覚されるのかについて検討を行った。その結果、アメリカ英語母語話者の正答率は他の母語話者に比べ全体的に高かった。日本語母語話者の正答率が最も低く、これは日本人被験者が語の意味と感情音声による情報の間に差異を感じたためであろうと述べている。この際には、「怒り」「喜び」「驚き」がよく混同され、その原因はピッチ曲線と声の大きさが類似していることであるとした。さらに、「悲しみ」と「疑い」も混同されるが、その原因はF0最小値が低いことと、声の大きさが小さいこととされた。また、韓国語母語話者は「怒り」について正答率が特に高いことも示された。

中林（2011）では、「韓国」「毎日」（頭高型）、「現金」「片仮名」（中高型）、「カラオケ」「温泉」（平板型）の6つの4拍語を刺激語とし、問い返し疑問文に含まれた「嫌」と「驚き」の感情について、ロシア人日本語学習者を対象に分析を行った。その結果、「嫌」という感情の認識が困難であることがわかった。また、アクセント核の有無によって感情の聴取の難易度が異なる可能性、さらに正答率の低い学習者では特定のアクセント型からの感情の聴取が困難である可能性が示唆された。

このように非母語話者が日本語音声から感情を知覚する際、母語話者と同様の傾向も認められるが、聞き手の母語が日本語感情音声の知覚に影響する可能性も示された。

3. 日本語母語話者による判定実験

3.1 刺激音の収録

多くの先行研究においては、エクマンが提案した基本六感情（Ekman 1992）が用いられている。本研究においては他の感情とよく共起するとされる「驚き」と「恐れ」を研究対象から除外し、通言語的に知覚されると報告されている「平静」（Calm）を追加した（Tickle 2000）。これにより、「喜び」「怒り」「悲しみ」「平静」の4感情を対象とした。

刺激語は、重野（2001）を参考に、単語自体が何らかの感情情報を示さず複数の感情を表

現しやすいとされる固有名詞と一般動詞、「小野さん」「ドイツ」「12 時」「分かった」を選択した。

これらの刺激語を 4 名の日本語母語話者（男女各 2 名、以下 JN）に、できるだけ自然かつ明確に感情を表出するように指示し、各感情とも 1 文につき 3 回発話してもらった。1 回目と 3 回目を除き、残り第 2 回目の録音を刺激音とした。このように、計 64 トークン（4 語×4 感情×4 名）の刺激音を収録した。

3.2 刺激音の評価と選択

刺激音の妥当性を評価するため、収録した 64 トークンをランダムに並べ替え、5 名の JN に聞かせ、4 つの感情から 1 つだけ選ぶように指示した。その結果、母語話者によってそれぞれ高い正答率が得られたため（「喜び」70.0%、「怒り」76.3%、「悲しみ」68.8%、「平静」88.8%）、64 トークン全てを聴取実験の刺激音とした。

4. 中国人日本語学習者を対象とした聴取実験

4.1 聴取実験

3.2 で得られた刺激音を用いて聴取実験を行った。被験者は中国語を母語とし、日本語を専攻する、日本滞在経験のない大学 1 年生 14 名、2 年生 14 名（日本語能力試験 N3 合格者）、3 年生 14 名（N2 合格者）および院生 14 名（N1 合格者）であり、合計 56 名であった。64 トークンを 16 例ずつランダムに並べ替え、4 試行を準備した。それぞれの刺激音提示の前に 260ms 間の長さの合図音を提示し、その 600ms 後に刺激音を提示した。刺激音の提示後、3s 間の解答記入時間を設けた。引き続き次の刺激音を提示した。試行間に休憩時間 15s を入れ、4 試行すべて終了するのに約 6 分間がかかった。

4.2 実験結果：中国人日本語学習者の知覚

表 1 に学年別の知覚結果を示す。Scherer et al (2001)では 9 ヵ国の母語話者を対象としたドイツ語の聴取実験を行い、9 ヵ国の母語話者がチャンスレベル以上の正答率で感情を正しく知覚できたと報告している。本研究の結果では、1、2 年生の正答率は 45.7%、55.3%程度であり、中国人初級学習者では日本語感情音声を正確に知覚できていない可能性が示された。各感情の知覚の傾向をより分かりやすく示すために、図 1 に JFL の平均正答率を示す。Scherer et al (2001)の指摘と同様に、「平静」の知覚が最も正答率が高く、次に「怒り」「悲しみ」という順に正答率が下がった。また、エリクソン・昇地（2006）に指摘されたように、「喜び」の正答率が最も低くなることが分かった。さらに、図 1 に示すように、1 年生から院生まで外国語の習得がすすむにつれ、正答率が上がることも示唆され

た。

表 1　各学年の感情知覚 confusion matrix（%）

1 年生	喜び	怒り	悲しみ	平静	2 年生	喜び	怒り	悲しみ	平静
喜び	**36**	27	16	21	喜び	**45**	27	5	22
怒り	17	**51**	15	17	怒り	19	**54**	13	14
悲しみ	8	15	**44**	33	悲しみ	5	8	**52**	35
平静	18	7	22	**53**	平静	9	6	16	**69**
3 年生	**喜び**	**怒り**	**悲しみ**	**平静**	**院生**	**喜び**	**怒り**	**悲しみ**	**平静**
喜び	**55**	24	6	15	喜び	**59**	29	5	7
怒り	12	**69**	7	12	怒り	17	**69**	8	6
悲しみ	2	3	**64**	31	悲しみ	1	3	**65**	31
平静	1	6	12	**81**	平静	1	2	14	**83**

表 1 の縦欄は刺激語の感情、横欄は回答された感情を示す。

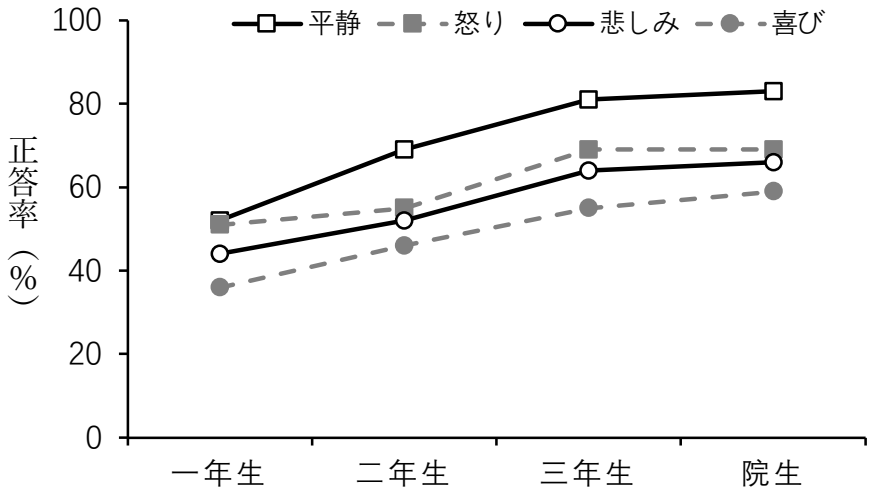


図 1　平均正答率

4.3 日本語感情音声の韻律的特徴

表 2 に「小野さん」のトークンを対象に、平均基本周波数(F0)、平均最高基本周波数(F0max)、平均最低基本周波数(F0min)、平均ピッチレンジ(F0range)、平均持続時間(Dur)およびインテンシティの平均最大値(Int)を示す。

表 2　「小野さん」の音響的特徴

	F0(Hz)	F0max(Hz)	F0min(Hz)	F0range(Hz)	Dur(ms)	Int(dB)
喜び	436	439	297	142	672	83
怒り	338	345	196	149	660	82
悲しみ	237	261	212	49	623	73
平静	289	305	234	71	666	80

「喜び」の発話においては、どの学年の学習者も平均 27%の高い確率で「怒り」との混同が観察された。これは、日本語の「喜び」の音響特徴が、F0 最高値が高く、発話持続時間が長く、インテンシティが大きいことに起因していると考えられる。これは、中国語の「怒り」（劉 2011）に似ており、このことが混同の原因であると考えられた。日本語の「怒り」は、ピッチレンジが最も大きく、ピッチ曲線の下降が顕著であり、声質特徴としてきしみ声（creaky voice）も観察された。エリクソン・昇地（2006）によれば、「怒り」は持続時間が一番短く、正答率も一番低いが、本研究の結果とは異なっていた。このことは「怒り」の度合いと関係があるのではないかと考えられる。靍・武田（2007）において、「激怒」の持続時間が「軽い怒り」より長いと述べられている。よって、感情の度合いも韻律的特徴に反映し、さらに知覚に影響を与えると考えられる。

また、「平静」と「悲しみ」も混同が見られたが、この原因は、低い基本周波数と幅の狭いピッチレンジという特徴が類似しているためと考えられる。「悲しみ」の音声は、平坦なピッチ曲線が特徴であり、F0 最高値と最低値も「平静」より低く、「平静」と区別している可能性が示唆された。

5. まとめ

本研究では、日本語母語話者が「喜び」「怒り」「悲しみ」「平静」の感情を込めた発話に対して、中国人日本語学習者がどのように知覚するかについて検討した。学習歴が浅い被験者の場合には、感情知覚の正答率が必ずしも高くはないが、学習時間の増加につれ、正答率も上がっていく可能性が示唆された。また、4 つの感情音声の中では「平静」の正答率が一番高く、「喜び」ではの知覚が最も困難であることが観察された。

次に、感情音声の音響的特徴についても分析を行った。「喜び」では、平均基本周波数、発話持続時間およびインテンシティともに一番高く、「悲しみ」では平均基本周波数、発話持続時間、インテンシティ、ピッチレンジともに一番低かった。さらに、「悲しみ」では、

平坦なピッチ曲線が観察され、低い基本周波数、狭いピッチレンジを示す「平静」と混同されることも観察された。今後は感情の種類を増やしてさらに検討していくことで、日本語学習者が日本語感情音声を知覚・産出する際に何が問題となるのかを明らかにしていきたい。

謝辞

本研究は科学研究費（基盤 B）課題番号 17H02352 の研究成果の一部である。

引用文献

Campbell, N. & Erickson, D. (2004) “What do People hear?A Study of the Perception of Non-verbal Affective Information in Conversational Speech,” Journal of the Phonetic Society of Japan 8(1), 9-28.

Ekman, P. (1992) “An Argument for Basic Emotions,” Cogniton And Emotion, 1992,6(3/4), 169-200.

Johnstone, T. (2001) “The effect of emotion on voice production and speech acoustics,” PhD Thesis, the University of Western Australia.

Scherer, K., Banse, R., Wallbott, H. (2001) “Emotion Inferences From Vocal Expression Correlate Across Languages and Cultures,” Journal of Cross-Cultural Psychology 32(1), 76-92.

Tickle, A. (2000) “English and Japanese Speakers’ Emotion Vocalization and Recognition: A Comparison Highlighting Vowel Quality, ” ISCA Workshop on Speech and Emotion. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.409.4169 (2017 年 7 月 30 日閲覧)

エリクソン ドナ ・昇地崇明(2006)「性差、および母語が感情音声の知覚に与える影響―日本語、韓国語、英語母語話者を対象として―」音声文法研究会（編）『音声文法』東京:くろしお出版, 31-46.

重野純(2001)「音声に含まれた感情の認知」青山心理学研究 1, 37-43.

田川恭識(2001)「『平静の問いかけ』と『非難の問いかけ』の弁別に対する F0 パタンの影響」聴覚研究会資料 31(10), 39-55.

龝真紀子・武田昌一(2007)「アナウンサーが発声した『怒り』の韻律的特徴と聴覚的印象の主成分分析を用いた比較」久留美信愛女学院短期大学研究紀要 30, 65-70.

中林律子(2011)「日本語音声に表れる感情の知覚―ロシア人日本語学習者の知覚の個人差に着目して―」音声研究 15(3), 14-25.

西端千香子(1996)「感情・態度を持つ音声の知覚・表出訓練のための CAI 用テキストの作成―促音が含まれる発話を中心に―」感情・態度を表す日本語音声の表出診断・訓練プログラムの構築に関する研究, 平成 7 年度科学研究費補助金, 一般研究(B)研究成果報告書, 21-40.

劉艳(2011)《普通话的情感语音韵律分析》, 南京师范大学硕士学位论文.

P02

日本語母語話者による英語音声の音節とストレス位置知覚に及ぼす諸要因

江口 小夜子（国際電気通信基礎技術研究所(ATR)，神戸大学大学院）
山田　玲子（国際電気通信基礎技術研究所(ATR)，神戸大学）

1. はじめに

日本語母語話者は英語音声の知覚において，英語と日本語の音声学的な違いから韻律の知覚が困難な場合があり，様々な観点からの研究が報告されている。例えば，Beckman (1986) や江口 (2015）は，英単語音声のストレス位置を判断させた結果，英語母語話者は，音の高さ，長さ，強さ，母音音質の 4 つの音響的特徴を手がかりとして区別するのに対し，日本語母語話者は，主に高さを手がかりにすることを示した。また，Tajima and Akahane-Yamada (2004)は，日本語母語話者に英単語の音声を聴覚呈示し，単語の音節数を数えさせた結果，音節数のカウントが困難であることを報告した。一方，韻律知覚には，音節数，子音の数，位置，その他様々な要因が影響している可能性があり，例えば Tajima and Akahane-Yamada (2004)では，音節構造の複雑さや語内の子音の数に影響して正答率が下がることが示されている。しかし，各要因の影響の度合いや差異は明らかになっていない。そこで本研究では，日本語母語話者を対象に，音節およびストレス位置の知覚実験を行い，子音構成，重子音の位置，音節数，ストレスの位置，母音構成，黙字，音節主音的子音の影響を系統的に比較した。本稿では，調査した要因のうち，影響の度合いが大きかった子音構成，重子音の位置，音節数，ストレスの位置を中心に報告する。

2. 方法

2.1. 実験参加者

日本語母語話者 17 名(男性 11 名，女性 6 名；19 歳～40 歳，平均年齢 24 歳)を実験参加者とした。実験参加者の TOEIC スコアは 330 点から 940 点に分布しており，英語習熟度の幅は広かった。アンケートにより全員 1 年以上の海外滞在経験がないこと，聴力や言葉の障害がないことを確認した。

2.2. 課題

Syllable Count Task と Stress Identification Task の 2 種類の課題を用いた。Syllable Count Task では,聞こえた英単語の音節数を 1～12 の選択肢ボタンから回答した。Stress Identification Task では，ストレスの位置を英単語の綴り（アルファベット）上をマウスクリックして回答した。刺激の呈示，反応の取得は PC 上の実験プログラムで制御した。刺激音の呈示はヘッドフォンを用いた。実験参加者にはできるだけ速く回答するように教示した。

2.3. 刺激

①子音構成，②重子音の位置，③音節数，④ストレス位置，⑤母音構成，⑥黙字，

⑦音節主音的子音の要因について調べるため刺激語を選定し，6 つの刺激セット(合計 1008 語)を作成した(表 1)。これらの語をアメリカ英語母語話者 4 名(男性 2 名，女性 2 名)が発音したものを刺激音とした。Syllable Count Task では, 1008 語(刺激セット 1~6)を，Stress Identification Task では，717 語(刺激セット 1，3，4)を使用した。それぞれランダムに 3 つのブロックに分けたため, ブロック A～F の合計 6 ブロックとなった。表 2 にブロックと課題およびトライアル数の関係を示す。

表 1. 刺激セットと調査した要因，語数の関係

刺激セット	調査する要因	語数	SCT	SIT
1	①子音構成	360	✔	✔
2	②重子音の位置	63	✔	
3	③音節数	213	✔	✔
	④ストレス位置		✔	✔
4	⑤母音構成	144	✔	✔
5	⑥黙字	108	✔	
6	⑦音節主音的子音	120	✔	

※SCT = Syllable Count Task で使用した刺激セットに✔マーク
SIT = Stress Identification Task で使用した刺激セットに✔マーク

表 2. 各ブロックで使用した課題とトライアル数

ブロック	課題	トライアル数
A	Syllable Count	336
B	Syllable Count	336
C	Syllable Count	336
D	Stress Identification	239
E	Stress Identification	239
F	Stress Identification	239

2.4. 手続き

6 つのブロック（A～F）を実験被験者毎にランダムな順序で実施した。また，各ブロックの問題呈示順序も実験参加者毎にランダムに出題した。

3. 結果

3.1. ①子音構成の影響

刺激セット 1 は 4 種類の子音構成(Syllable Complexity)と 3 種類の音節数(1σ～3σ; 本稿では, 刺激語を構成する音節数を σ で示す)を組み合わせた条件を満たす 252 語で構成された。子音構成条件は下記のとおり。

・SC0：母音の前後に単独（連鎖しない）子音がある　CVC, CVCVC, CVCVCVC
・SC+1：2 つの子音連鎖を含む　CCVC, CVCC, CCVCVC, CVCVCC, etc.
・SC+2：3 つの子音連鎖を含む　CCCVC, CVCCC, CCCVCVC, CVCVCCC, etc.

3.1.2. Syllable Count Task の結果

各実験参加者の刺激条件毎の正答率を従属変数とし，子音構成(SC0，SC+1，SC+2)と音節数(1σ～3σ)を要因とした 2 要因の分散分析を行った。その結果，子音構成および音節数の主効果はなく，交互作用[$F(4,64)=2.763$, $p<.05$]が有意だった(図 1)。下位検定の結果も合わせてまとめると，1 音節語のみ，SC0, SC+1 > SC+2 の順に正答率が低かった。つまり，3 重子音の刺激に対して，正答率が低くなることが示唆された。

次に，回答と正解の差分（「回答した音節数」－「正解音節数」）を従属変数とし，同様の分散分析を行った結果，子音構成の主効果[$F(2,32)=17.460$, $p<.001$]，および音節数の主効果[$F(2,32)=22.578$, $p<.001$]が有意だった(図 2)。下位検定の結果を合わせてまとめると，回答と正解の差分は，SC0 < SC+1 < SC+2 の順で，プラスの方向へ有意に

大きくなった。したがって，子音構成が複雑なほど，実際の音節数より多く数えることが示されたと言える。また，3σ < 2σ < 1σ の順で，プラスの方向へ有意に大きくなった。つまり，音節数が少ないほど，実際の音節数より多く数えることも示された。

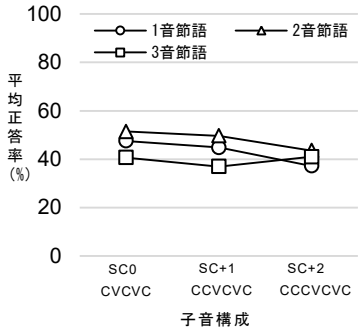


図 1: Syllable Count Task における刺激条件毎の正答率

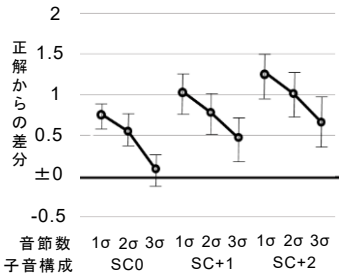


図 2: Syllable Count Task における正解と回答の差分

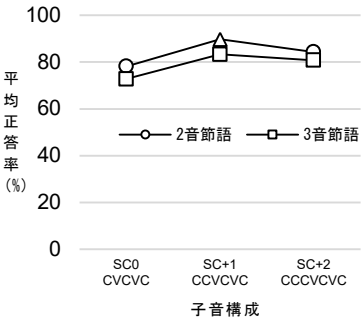


図 3: Stress Identification Task における刺激条件毎の正答率

3.1.3. Stress Identification Task の結果

各実験参加者の刺激条件毎の正答率を従属変数とし，子音構成(SC0，SC+1，SC+2)と音節数(2σ,3σ)を要因とした 2 要因分散分析を行った。その結果，交互作用はなく，子音構成の主効果[F(2,32)=8.846, p<.001]，音節数の主効果[F(1,16)=5.226, p<.05]が有意だった(図 3)。下位検定の結果を合わせてまとめると，SC0 の正答率が他の条件より有意に低かった。つまり，子音構成が複雑になると正答率が高くなることが示唆された。

3.2. ②重子音の位置の影響

3.2.1. Syllable Count Task の結果

刺激セット 1，2 から，2 種類の重子音(2 重子音，3 重子音)，その出現位置(語頭，語末)および 3 種類の音節数(1σ～3σ)を組み合わせた条件を満たす 144 語を抽出して用いた。

3.2.1. Syllable Count Task の結果

各実験参加者の刺激条件毎の正答率を従属変数とし，重子音(2 重子音，3 重子音)，位置(語頭，語末)，音節数(1σ～3σ)を要因とした 3 要因の分散分析を行った。その結果，主効果および 2 次の交互作用はなく，単純交互作用が全て有意だった(重子音と位置[F(1,16)=7.696, p<.05]，重子音と音節数[F(2,32)=5.257, p<.05]，重子音の位置と音節数[F(2,32)=6.845, p<.005])。下位検定の結果を合わせてまとめると，1 音節語は，2 重子音，3 重子音ともに語頭の正答率が語末よりも有意に低かったが，他の条件については統一的な傾向は認められなかった。

次に，回答と正解の差分（「回答した音節数」－「正解音節数」）を従属変数とし，分散分析を行った結果，主効果，単純交互作用，2 次の交互作用が全て有意だった(重子音[F(1,16)=19.042, p<.001], 位置[F(1,16)=26.862, p<.001], 音節数[F(2,32)=37.996, p<.001], 重子音と位置[F(1,16)=9.495, p<.01]，重子音と音節数[F(2,32)=4.048, p<.05]，重子音と位置と音節数[F(2,32)=13.182, p<.001])(図 4，5)。下位検定の結果を合わせてまとめると，正解からの差分は，2 重子音と 3 重子音ともに，語頭 > 語末の順，また，1σ > 2σ > 3σ の順で低かった。したがって，重子音は語末よりも語頭にある場合，および音節数が

少ない場合に，実際の音節数より多く数えることが示唆された。

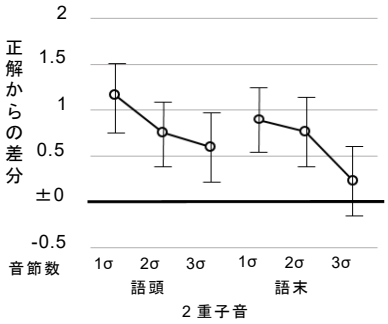


図 4: Syllable Count Task における
2 重子音の正解と回答の差分

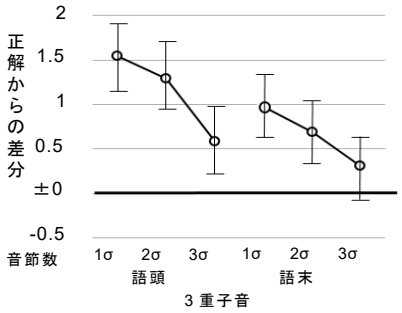


図 5: Syllable Count Task における
3 重子音の正解と回答の差分

3.3. ③音節数の影響

刺激セット 3 は音節数(1σ～6σ)の条件を満たす 213 語で構成された。

3.3.1. Syllable Count Task の結果

各実験参加者の刺激条件毎の正答率を従属変数とした F 検定を行ったところ，音節数(1σ～6σ)の効果は有意だった[F(5,80)=6.568, p<.001] (図 6)。多重比較の結果をまとめると、2σ > 1σ，3σ > 4σ > 5σ > 6σ の順で正答率が低くなった。したがって，1 音節語以外の条件では，音節数が増えるにつれ，正答率が低くなることが示唆された。

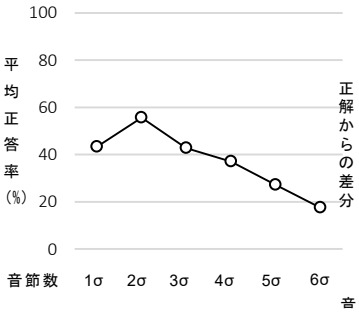


図 6: Syllable Count Task に
おける刺激条件毎の正答率

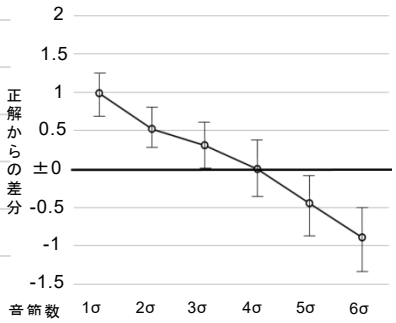


図 7: Syllable Count Task に
おける正解と回答の差分

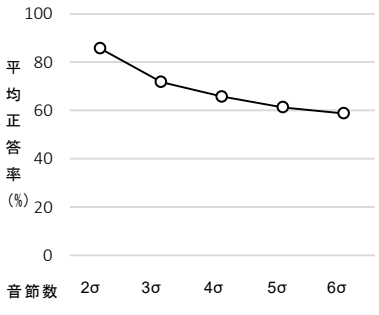


図 8: Syllable Count Task に
おける刺激条件毎の正答率

次に，回答と正解の差分（「回答した音節数」－「正解音節数」）を従属変数とした F 検定を行った結果，音節数の効果は有意だった[F(5,80)=39.994, p<.001] (図 7)。多重比較の結果をまとめると，6σ < 5σ < 4σ < 3σ < 2σ < 1σ の順で，正解からの差分はプラスの方向に大きくなった。また，1σ～3σ の差分はプラスであったが、5σ，6σ の差分はマイナスであった。つまり，音節数が少ない 1～3 音節語では，実際の音節数より多く数えるのに対し，音節数の多い 5～6 音節語の音節数は実際の音節数より少なく数えることが示唆された。

3.3.2. Stress Identification Task の結果

各実験参加者の刺激条件毎の正答率を従属変数とし，音節数(2σ～6σ)を要因とした F 検定を行った。その結果，音節数の効果が有意だった[F(4,64)=30.523, p<.001] (図 8)。多重比較の結果をまとめると，2σ > 3σ > 4σ > 5σ > 6σ の順で正答率が低くなった。

したがって，音節数が増えるにつれ，正答率が低くなることが示唆された。

3.4. ④ストレス位置の影響

刺激セット 3 から，2〜6 音節の 201 語に対する結果を分析した。刺激条件はストレスの位置であった。音節数ごとに刺激条件の水準数は異なっており，2 音節語のストレス位置条件は，1 音節目と 2 音節目の 2 水準，3 音節語のストレス位置条件は 1 音節目，2 音節目，3 音節目の 3 水準であった。6 音節語のみ，1 音節目と 6 音節目にストレスがある単語の数が不十分であったため，それらを除外した 4 水準であった。

3.4.1. Syllable Count Task の結果

各実験参加者の刺激条件毎の正答率を従属変数とし，刺激条件(2σ〜6σ)ごとにストレス位置を要因とした F 検定を行ったが，有意差は認められなかったことから，ストレス位置の影響がないことが示唆された。

次に，回答と正解の差分（「回答した音節数」−「正解音節数」）を従属変数として，刺激条件(2σ〜6σ)ごとにストレス位置を要因とした F 検定を行った結果，2，4，5，6 音節語においてストレス位置の効果が有意であった(2σ[F(1,16)=9.781, p<.01]; 4σ[F(3,48)=9.127, p<.001]; 5σ[F(4,64)=9.613, p<.001]; 6σ[F(3,48)=3.994, p<.05])。多重比較の結果をまとめると，ストレスの位置が語末位置に近づくほど，正解からの差分がプラスの方向へ大きくなることが示された。

3.4.2. Stress Identification Task の結果

各実験参加者の刺激条件毎の正答率を従属変数とし，刺激条件(2σ〜6σ)ごとにストレス位置を要因とした F 検定を行った。その結果，全ての条件においてストレス位置の効果が有意だった(2σ[F(1,16)=11.620, p<.005]; 3σ[F(2,32)=26.847, p<.001]; 4σ[F(3,48)=26.176, p<.001]; 5σ[F(4,64)=19.087, p<.001]; 6σ[F(3,48)=4.087, p<.05])(図 9)。多重比較の結果をまとめると，ストレスの位置が語末位置に近づくほど，正答率が有意に低くなることが示唆された。

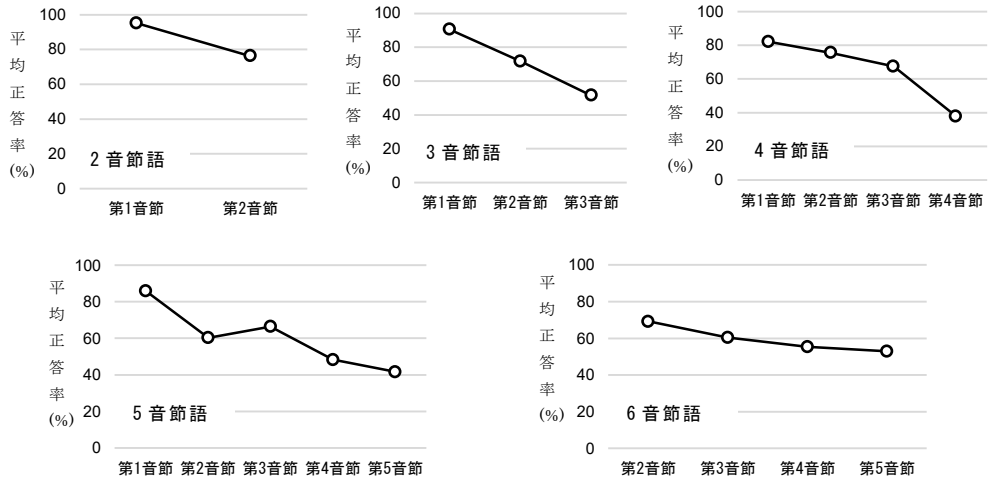


図 9. Stress Identification Task における
2 - 6 音節語の刺激条件毎の正答率

4. 考察

日本語母語話者の全刺激音に対する正答率は，音節知覚では 39.3%，ストレス位置知覚では 69.5%とともに低かった。課題が異なるため両知覚の難易度の直接の比較はできないものの，日本語母語話者は，ストレス知覚のみならず音節数のカウントも困難であることが示唆された。

表 3. 音節知覚およびストレス位置知覚における各要因の影響

	音節知覚	ストレス位置知覚
子音構成	1 音節語の 3 重子音で困難	子音構成が複雑な場合に正答率が高い
重子音の位置	1 音節語の 2 重子音，3 重子音において，重子音が語頭にある場合に困難	
音節数	音節数が多くなるほど困難	音節数が多くなるほど困難
ストレス位置	影響なし	ストレス位置が語末位置に近づくほど困難

各要因の影響について，まとめた結果を表 3 に示す。音節知覚では，日本語母語話者は，音節数が多くなるほど，音節知覚が困難になることが分かった。また，1 音節語の刺激において，子音構成および重子音の位置の影響が認められたが，その他の条件では影響がなかったことから，子音構成，重子音の位置の影響の度合いは比較的小さいと考えられる。次に，ストレス位置知覚では，母音の前後に単独子音がある条件の正答率は低く，2 重子音，3 重子音など子音構成が複雑な場合に正答率が高くなることが示された。また，音節数が多くなるほど，およびストレス位置が語末位置に近いほど知覚が困難になることが分かった。子音構成やストレス位置が影響する理由について明らかにするためには，今後さらなる検討が必要である。正解との差分のデータからは，調査した要因全てにおいて，音節数が少ないほど，実際の音節数より多く数えることが分かった。

以上の結果から，総じて，今回対象とした要因は複雑に音節知覚およびストレス位置知覚に影響していることが示された。今後は，本研究の結果から明らかになった要因に焦点をあて，韻律の学習方法について検討したい。

参考文献

Beckman, M. E. (1986) *Stress and non-stress accent.* Dordrecht: Foris.

江口小夜子 (2015)「英語音声のストレス知覚における母音音質の影響 -英語母語話者と日本語母語話者の比較-」『第 29 回日本音声学会全国大会予稿集』68-72.

Tajima, K. and Akahane-Yamada, R. (2004) “Production and perception of syllable structure in second-language speech.” *The 18th International Congress on Acoustics, Proc. ICA 2004, IV,* 3321-3324.

P03

中国語母語話者による日本語特殊拍の知覚 —特殊拍間の混同—

張 林姝（神戸大学大学院）林 良子（神戸大学）山田 玲子（国際電気通信基礎技術研究所＜ATR＞・神戸大学）

1. はじめに

日本語特殊拍に関する研究は多数報告されているが（皆川他 2002, 戸田 2003, Kubozono et al. 2011 等）、それらのほとんどは特殊拍と自立拍の混同を対象にしたものである。しかし、特殊拍に分類される 3 つの音素である長音/R/、撥音/N/、促音/Q/の間、即ち特殊拍間でも混同が起こる可能性がある。

岡田（2003）は、生成における/R/と/N/の混同について報告した。上野（2014）は、母語話者が/N/を/R/に（例えば、「原因」を「ゲーイン」に）発話する場合があると述べている。知覚に関して、黒崎（2002）は、「テーイン」を単独で聞くと正しく「テーイン」に聞き取る確率は 60%だが、「コンビニのテーインをしている」という文では、「テンイン」に聞き取る確率が高くなり、「テーイン」と聞き取る確率は 0%になると報告した。また、本橋（2005）は英語母語話者が促音を長音に聞き取る傾向にあることを示した。

しかし、これらの特殊拍間の混同を取り上げた先行研究では、限られた対の混同しか扱っておらず、使用された刺激語も限られている。本研究では、特殊拍の混同を系統的に検討するために、/R/、/N/、/Q/、自立拍<I>[注]の相互の関係について日本語母語話者と中国人日本語学習者を対象に 2 つの知覚実験を通して検討した。実験 1 では、特殊拍で対立する無意味語 2 つの同-異判断実験を行った。実験 2 では、文内の、特殊拍が含まれる無意味語を対象に、同定実験を行った。

2. 実験 1

2.1. 方法

2.1.1. 刺激音

4 種類の拍（R/N/Q/I）で対立する 3 音節 4 拍の無意味語/ta?paka/、/tapa?ka/、/tapaka?/（?=/R//N//Q/またはなし）を用いた。アクセント型 4 種類（平板/頭高/中高 I/中高 II）と特殊拍の生起位置 3 種類（第 1 音節/第 2 音節/第 3 音節）を組み合わせた 12 条件のうち、「ターパカ/タンパカ/タッパカ/タパカ」など 4 種類の拍の対立を実現できる 6 セットの 24 語を素材とした（表 1）。ただし、自立拍の語はセット間で重複があり、重複を除くと 21 語だった。

東京方言母語話者（女性、40 代）が防音室でこれらの語をそれぞれ 5 回読み上げたものを、IC レコーダーを用いて、サンプリング周波数 44.1kHz、量子化精度 16bit の音声ファイルとして保存した。各語 5 回の発話から時間長が中央値を示す発話を刺激音に使用した。刺激語の長さは、2 音節語では平均 365.95msec、3 音節語では平均 534.66msec だった。

同一セット中の異なる 2 語を組み合わせた 6 対（例えば、ターパカ-タンパカ/タンパカ-タッパカ/タッパカ-ターパカ/タパカ-ターパカ/タパカ-タンパカ/タパカ-タッパカ）を 2 種

注：本文では自立拍を<I>（Independent morae）と記す。

類の順序（AB/BA）で連結し 1 セットから 12 個（6 対×2 種類の順序）の刺激音を作成した。刺激語と刺激語の間には 1000msec の無音区間を挿入した。6 セット、合計 72 個（12 個×6 セット）の刺激音に加え、ダミーとして同じ語をつなげた刺激語対の刺激音を作成した。

2.1.2. 実験参加者

日本語母語話者 15 名（20～40 才）と中国人日本語学習者 15 名（20 才）が参加した。全員健聴者で、神戸大学の大学生と大学院生であった。中国人日本語学習者は全員日本語能力試験 N1 の合格者だった。

2.1.3. 手続き

静かな部屋でヘッドホンを通して刺激音をランダムに呈示した。実験参加者には、ヘッドホンから聞こえてくる対となる２つの単語が同じか否かを判断し、できる限り早くキーボードの割り当てられたキーを押して回答するように教示した。刺激の呈示と反応の取得は SuperLab（ver.5）を用いた。

表 1 実験 1 の刺激語

分析に用いた語を太枠で示した。日本語の音韻規則に沿わない語は背景に色を付けて示した。

拍の種別	平板型			頭高型			中高型（核が 2 拍目）			中高型（核が 3 拍目）		
	1*	2	3	1	2	3	1	2	3	1	2	3
R	ターパカ―	タパーカ―	タパカー―	タ\ーパカ	タパー\カ	タ\パカー	ター\パカ	タパ\ーカ	タパ\カー	ターパ\カ	タパー\カ	タパカ\ー
N	タンパカ―	タパンカ―	タパカン―	タ\ンパカ	タ\パンカ	タ\パカン	タン\パカ	タパ\ンカ	タパ\カン	タンパ\カ	タパン\カ	タパカ\ン
Q	タッパカ―	タパッカ―	タパカッ―	タ\ッパカ	タ\パーカ	タ\パカッ	タッ\パカ	タパ\ッカ	タパ\カッ	タッパ\カ	タパッ\カ	タパカ\ッ
I	タパカ―	タパカ―	タパカ―	タ\パカ	タ\パカ	タ\パカ	タパ\カ	タパ\カ	タパ\カ	タパ\カ	タパ\カ	タパ\カ

*特殊拍が何音節目にあるかを示す。

2.2 結果

正答率は、日本語母語話者は 98.3%、中国人日本語学習者は 96.9%と高く、グループ間の差がなかった。そこで反応時間を従属変数とし、母語（日本語母語話者/中国人日本語学習者）、アクセント型（平板型/中高型/頭高型）、対の種別（「特殊拍-特殊拍」対/「特殊拍-自立拍」対）を要因とした 3 元配置分散分析を行った。その結果、母語とアクセント型の交互作用[$F(2, 116)=4.16$, $p<.05$]、アクセント型と対の種別の交互作用[$F(2, 116)=18.34$, $p<.001$]、母語の主効果[$F(1, 58)=18.32$, $p<.05$]が有意だった。中国人日本語学習者は日本語母語話者より反応時間が長く、下位検定の結果、日本語母語話者で平板型と中高型に差があったが、中国人日本語学習者はアクセント型による差がなかった（図 1）。また、「特殊拍-特殊拍」対と「特殊拍-自立拍」対の差が中高型のみで現れた。以下、アクセント型ごとに、3 元配置分散分析（母語・生起位置・対の種別）を行った。

平板型では、3 つの要因の交互作用がなく、母語の主効果のみ有意で[$F(1, 28)=9.74$, $p<.005$]、中国人日本語学習者の反応時間は日本語母語話者より長かった（図 2）。対の種別について、交互作用、主効果ともになかったことから、特殊拍同士の弁別の反応時間は、特殊拍と自立拍の弁別の場合と差がなかったといえる。頭高型では、母語と対の種

別の交互作用 [F(1, 28)=5.86, p<.05]、母語の主効果[F(2, 116)=4.16, p<.05]、生起位置の主効果[F(1, 28)=11.30, p<.005]が有意だった（図 3）。日本語母語話者の「特殊拍-特殊拍」対に対する反応時間は「特殊拍-自立拍」対より長かったが、中国人日本語学習者では、差がなかった。中国人日本語学習者の反応時間は拍の種別にかかわらず、日本語母語話者より長かった。特殊拍が第 2 音節にある際の反応時間は第 1 音節の場合より長かった。

頭高型では、母語と対の種別の交互作用 [F(1, 28)=5.86, p<.05]、母語の主効果[F(2, 116)=4.16, p<.05]、生起位置の主効果[F(1, 28)=11.30, p<.005]が有意だった（図 3）。日本語母語話者の「特殊拍-特殊拍」対に対する反応時間は「特殊拍-自立拍」対より長かったが、中国人日本語学習者では、差がなかった。中国人日本語学習者の反応時間は拍の種別にかかわらず、日本語母語話者より長かった。特殊拍が第 2 音節にある際の反応時間は第 1 音節の場合より長かった。

中高型では、生起位置の要因とアクセント核の位置の要因が共起しているが、交互作用はどの要因の間にもなく、母語 [F(1, 28)=7.70, p<.01]、生起位置 [F(1, 28)=8.12, p<.01]、対の種別 [F(1, 28)=24.93, p<.001]、すべての主効果が有意だった（図 4）。中国人日本語学習者の反応時間が日本語母語話者より長かった。第 2 音節の反応時間が第 1 音節より長かった。また、「特殊拍-自立拍」対の反応時間が「特殊拍-特殊拍」より長かった。

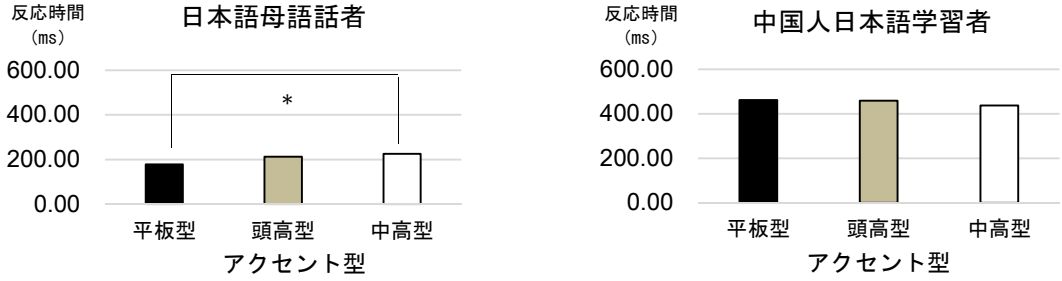


図 1 実験 1:弁別課題におけるアクセント型と反応時間の関係。

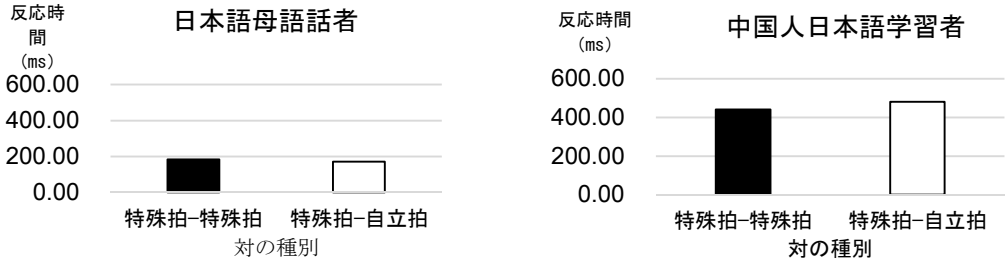


図 2 実験 1:平板型における対の種別と反応時間の関係。

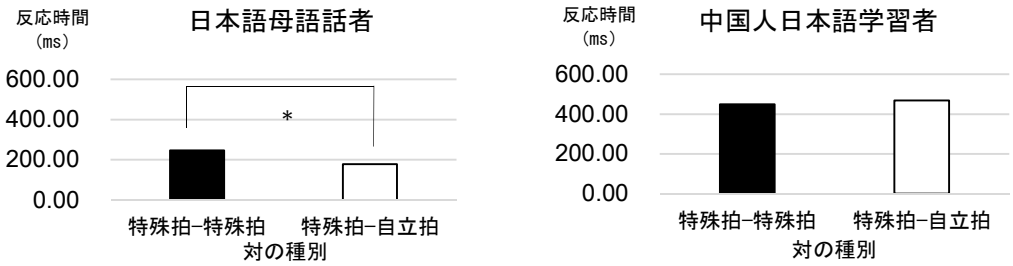


図 3 実験 1:頭高型における対の種別と反応時間の関係。

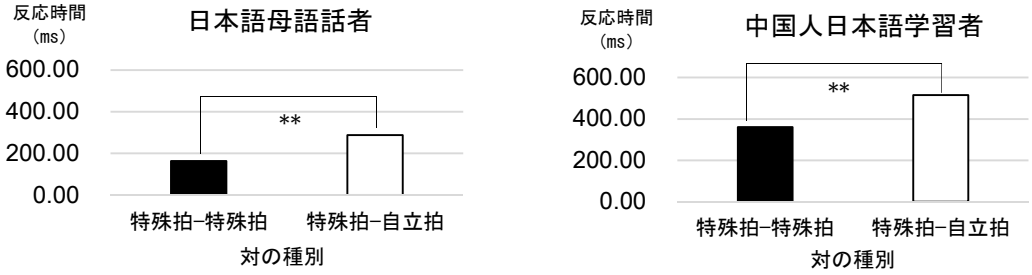


図 4 実験 1:中高型における対の種別と反応時間の関係

以上のすべての分析で、母語の主効果が有意だったことから、拍で対立する 2 語の弁別の反応時間は、中国人日本語学習者は日本語母語話者より長いことが示された。次に、拍の生起位置の主効果が有意だったことから、特殊拍が第 2 音節に生起する場合は第 1 音節に生起する場合より長いことが分かった。また、対の種別については、反応時間が平板型では「特殊拍-特殊拍」＝「特殊拍-自立拍」、頭高型では「特殊拍-特殊拍」＞「特殊拍-自立拍」、中高型では「特殊拍-特殊拍」＜「特殊拍-自立拍」だったことから、「特殊拍-特殊拍」は「特殊拍-自立拍」より知覚困難である可能性が示唆された。

3. 実験 2

3.1. 方法

3.1.1. 刺激音

拍 4 種類（R/N/Q/I）を 1 セットにし、3 音節 4 拍の無意味語を作成した。特殊拍の前後に必ず/ka/と/ta/を入れ、特殊拍が第 1（語頭）、2（語中）、3（語尾）のいずれかの音節に現れるようにした。アクセント型は全て平板型とした。刺激語は合計 11 語となった（表 2）。

これらの刺激語をキャリア文の中に入れ、刺激文とした。キャリア文は、刺激語が外来語と解釈されるもの、例えば「去年オーストラリアで新型ロボット　＿＿＿が開発されました。」を、5 種類作成した。1 つの語を 2 種類のキャリア文に入れたので、刺激文は合計 22 文となり、ダミー語の入ったダミー文 16 文を加え、合計 38 文を用いた。全 38 文を東京母語話者 1 名（女性、40 代）が防音室で読みあげたものを、IC レコーダーを用いて録音した。なるべく自然に発話するように教示した。サンプリング周波数 44.1kHz、16bit の精度でリニア wav ファイルとして保存した。

3.1.2. 実験参加者

日本語母語話者 15 名（20～40 才）と中国人日本語学習者 15 名（20 才）が参加した。全員健聴者で、神戸大学の大学生と大学院生であった。中国人日本語学習者は全員日本語能力試験 N1 の合格者だった。

3.1.3. 手続き

実験参加者は静かな部屋で、ヘッドホンを通して刺激文を聴取し、刺激文の再生終了 3 秒後に画面上に呈示される 4 つの選択肢から、聞いた刺激文の中に入っている無意味語と同じものを選び、なるべく速くそれに対応するキーを押すように教示した。選択肢は刺激語とそれと対立する 3 語だった。刺激、選択肢呈示は、SuperLab （Ver. 5）を用いた。

表 2 実験 2 の刺激語

日本語の音韻規則に沿わない語は背景に色を付けて示したが、それらは刺激語に使わなかった。

拍の種別	第 1 音節（語頭）	第 2 音節（語中）	第 3 音節（語尾）
R	カータパ	パカータ	パタカー
N	カンタパ	パカンタ	パタカン
Q	カッタパ	パカッタ	パタカッ
I	カタパ	パカタ	パタカ

3.2. 結果

正答率を従属変数とし、母語と生起位置を要因とした分散分析を行なった結果、交互作用が有意で［F(3, 168)=5.375, p<.005］、中国人日本語学習者のみ、正答率が生起位置によって異なった（図 5）。生起位置ごとに、母語（日本語母語話者/中国人日本語学習者）と拍の種別（R/N/Q/I）を要因に配置分散分析した結果を以下に記す。

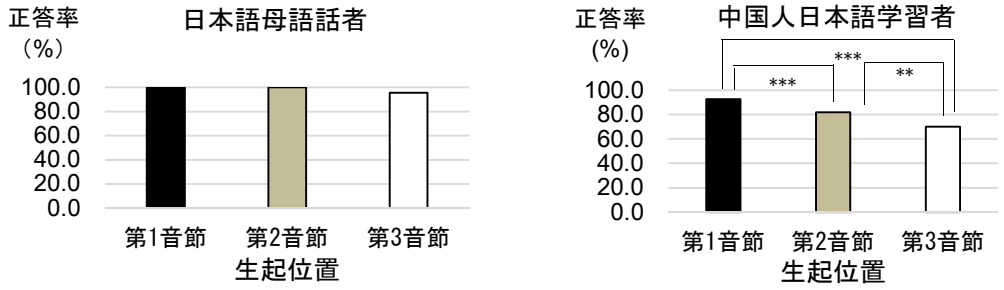


図 5 実験 2:同定課題における特殊拍の生起位置と正答率の関係。

第 1 音節では、交互作用がなく、母語の主効果のみ有意だった[F(1, 28)=4.295, p<.05]。中国人日本語学習者の正答率は日本語母語話者より有意に低かった。第 2 音節では、母語と拍の種別の交互作用 [F(3, 84)=3.961, p<.05]が有意だった。日本語母語話者は拍の種別に関わらず、ほぼ 100%の正答率を示したが、中国人日本語学習者は拍の種別による正答率の変動が大きく、/R/における正答率は他より低かった。また、日本語母語話者と中国人日本語学習者の差が/R/と/N/のみにあった。第 3 音節では、母語と拍の種別の交互作用[F(2, 116)=4.08,p<.05]が有意だった。日本語母語話者は拍の種別に関わらず、ほぼ 100%の正答率を示した。中国人学習者は拍の種別による正答率の変動が大きく、特殊拍の正答率は自立拍より低かった。日本語母語話者との差が特殊拍のみにあった。

回答パターンは図 6～8 に示す。誤答パターンを見ると、第 1 音節で、日本語母語話者も中国人日本語学習者も特殊拍間の混同はほぼなかった（図 6）。第 2 音節で、日本語母語話者には混同がなかったが、中国人日本語学習者には/R/と/Q/の互いの混同があった（図 7）。第 3 音節で、日本語母語話者には混同がなかったが、中国人日本語学習者には/N/から/R/への混同があった（図 8）。

日本語母語話者の同定率（%）

選択肢＼刺激語	R	N	Q	I
R	100.0	0.0	0.0	0.0
N	0.0	100.0	0.0	0.0
Q	0.0	0.0	100.0	0.0
I	0.0	0.0	0.0	100.0

中国人日本語学習者の同定率（%）

選択肢＼刺激語	R	N	Q	I
R	96.7	0.0	3.3	13.3
N	0.0	96.7	0.0	0.0
Q	0.0	0.0	90.0	0.0
I	3.3	3.3	6.7	86.7

図 6 実験 2:第 1 音節における拍の種別と回答パターンの関係。

日本語母語話者の同定率（%）

刺激語 / 選択肢	R	N	Q	I
R	100.0	0.0	0.0	0.0
N	0.0	100.0	0.0	0.0
Q	0.0	0.0	100.0	0.0
I	0.0	0.0	0.0	100.0

中国人日本語学習者の同定率（%）

刺激語 / 選択肢	R	N	Q	I
R	63.3	3.3	10.0	3.3
N	3.3	90.0	3.3	3.3
Q	20.0	0.0	83.3	0.0
I	13.3	6.7	3.3	93.3

図 7 実験 2:第 2 音節における拍の種別と回答パターンの関係。

日本語母語話者の同定率（%）

刺激語 / 選択肢	R	N	I
R	100.0	3.3	3.3
N	0.0	96.7	6.7
I	0.0	0.0	90.0

中国人日本語学習者の同定率（%）

刺激語 / 選択肢	R	N	I
R	80.0	16.7	23.3
N	0.0	53.3	0.0
I	20.0	30.0	76.7

図 8 実験 2:第 3 音節における拍の種別と回答パターンの関係。

以上の結果から、 拍の同定において、中国人日本語学習者は語尾に行くほど、特殊拍同定の正答率が下がり、特殊拍間の混同も起こることが分かった。

4. 考察

実験 1 の結果からは、弁別の際に「特殊拍-特殊拍」のほうが「特殊拍-自立拍」より反応時間が長い場合があることが示された。実験 2 の結果からは、同定時に、R と Q を相互に混同する、N を R と混同するなど、特殊拍間で混同される場合があることが示された。これらの結果から、中国語母語話者は特殊拍同士も混同していることが明らかになり、日本語の学習の過程で「特殊拍-自立拍」の混同のみでなく、特殊拍同士の混同を解決するための訓練も行う必要性が示唆された。より良い学習方法を提案するためには、今後、生成の面からも考察する必要がある。

参考文献

岡田祥平（2003）「撥音から長音への「言い間違い」現象について：『日本語話し言葉コーパス』を資料として（第 17 回全国大会発表要旨）」『音声研究』4:3, 117.

上野善道（2014）「フンイキ＞フインキの変化から音位転換について考える」『生活語の世界』8-19.

黒崎典子（2002）「母音に前接する撥音について：日本語母語話者にとっての知覚の難易」『神奈川大学言語研究』25, 11-22.

戸田貴子（2003）「外国人学習者の日本語特殊拍の習得」『音声研究』7:2, 70-83.

皆川泰代・前川喜久雄・桐谷滋（2002）「日本語学習者の長/短母音の識別におけるピッチ型と音節位置の効果」『音声研究』6:2, 88-97.

本橋美樹（2005）「英語話者による促音の認識」『言語文化と日本語教育』30, 95-98.

Kubozono, H., & Takeyasu, H., & Giriko, K., & Hirayama, M.（2011）" Pitch cues to the perception of consonant length in Japanese." *International Congress of Phonetic Sciences August2011*. (pp, 17-266).

P04

後続子音の調音方法の日本語話者による英語の母音の同定と弁別への影響

野澤　健（立命館大学）
t-nozawa@ec.ritsumei.ac.jp

1. はじめに

一般に、母語以外の言語の母音の知覚実験を行う場合、前後の子音の影響を最小限にするため/hVt/のような構造が用いられる。また、子音の影響を考慮する場合、先行する子音の調音位置の影響が中心であった(Nozawa & Wayland 2012 など)。

このような現状を踏まえて、母音の音響的特性に大きな影響を与えることは知られていても、それが非母語話者の母音の知覚にどのように影響するかはあまり研究されて来なかった後続子音の調音方法に焦点を当て、母音の知覚へのその影響ついて述べていく。具体的には、鼻音/n/と接近音/l/の影響である。鼻音は母音の F1 領域に鼻音のフォルマントを起こすため、母音の高さに関する情報が影響を受ける可能性がある。英語の母語話者を対象にした実験では、後続の鼻音により鼻音化する文脈(/bṼnd/)で鼻音化した母音の高さの知覚は影響されなかったが、本来鼻音化しない文脈での鼻音化した母音̃/bṼd/は、対応する口腔母音よりも低く知覚されるという報告がある(Krakow et al. 1988)が、アメリカ英語の一部の方言で鼻音の前で/ɛ/-/ɪ/中和する(pen/pin merger)こと(Labov et al. 2005)や音素として鼻母音を持つ言語では鼻母音の音素数は口腔母音の音素数よりも少ないこと(Beddor, 1993)などからも鼻音は母音を不明瞭にすると考えられる。また、アメリカ英語固有の現象として、鼻音の前で/æ/が緊張母音化(tensing)し、[eə]のような響きを持つことが挙げられる。図 1 に同じ話者の発話による pat と pan の波形とスペクトラムを示す。Pan の方が F2 が大きく曲がっているのがわかる。

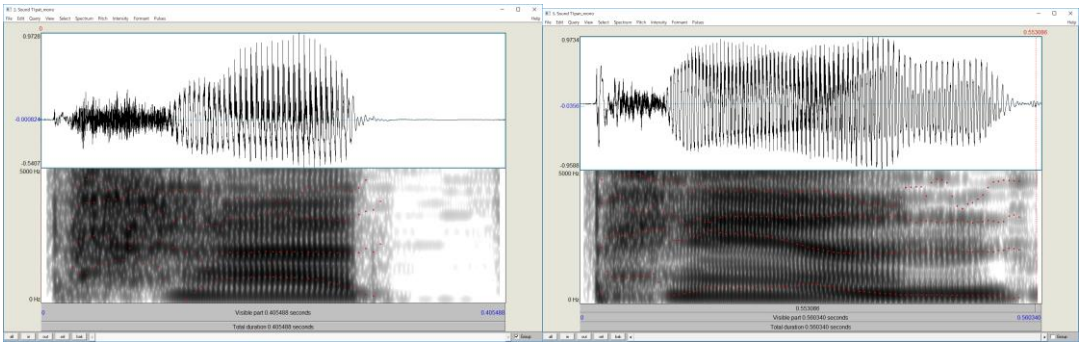


図 1;　アメリカ英語母語話者の発話による pat と pan の波形とスペクトログラム

一方、母音に後続する/l/は、先行する母音の F2 を継続的に降下させることが知られている(Olive et al. 1993)。図 2 に同じ話者の発話による Pete と peel の波形とスペクトラムを示す。Peel では、F2 が連続して下降しているのがわかる。

このように/n/や/l/の前では、母音が閉鎖音の前とは異なる動きを見せるので、非母語話者である日本語話者の英語の母音の同定・弁別に大きな影響を及ぼすのではないかと考えられる。

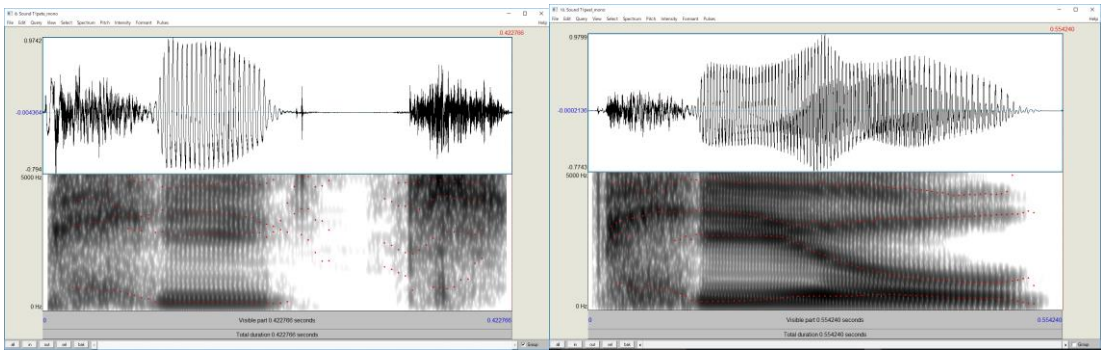


図 2：アメリカ英語母語話者の発話による Pete と peel の波形とスペクトログラム

2. 実験

2.1. 音声刺激

音声刺激は、これまでの実験(Nozawa & Wayland 2012 等)で使用した音声及び同時期に録音した音源を使用した。4 人のアメリカ英語の母語話者(女性)の発話をアラバマ州オーバーンでサンプル周波数 44.1KHz で録音した。4 人の話者の内訳は、ニューヨーク州出身 2 名、カリフォルニア州出身 1 名、ウィスコンシン州出身 1 名である。これらの話者が/ i, ɪ, ɛ, æ, ɑ, ʌ/の 6 母音を/hVt/, /pVt/, pVn/, /pVl/の子音環境の中で発したものを音声刺激として使用できるように編集したものを実験に使用した。6 母音×4 子音環境×4 話者=96 の音声刺激を使用することとなった。

2.2. 実験参加者

実験参加者は、日本の大学に通う日本語話者 33 名(34 名が参加したが、最後まで参加できなかったので、33 名のデータについて述べる)である。内訳は男性 5 名、女性 28 名で、実験参加時の平均年齢は 19.4(18-22)才である。英語力を示す指標としての TOEIC（R）の得点は、自己申告で 330 点から 990 点で平均 628 点(5 名が受験経験がない)である。英語以外の外国語の学習歴があると回答したものが 8 名、海外での生活経験があるものが 10 名、うち 1 名は海外(マレーシア)生まれであった。全員、聴覚は正常である。

2.3. 実験方法

実験参加者は、以下の３つの実験に参加した。実験用の音声はすべてコンピュータで再生され、実験参加者はヘッドフォンを通じて音声を聴いた。

2.3.1 知覚同化実験

実験参加者は、1 回の試行に 1 つの音声刺激を聴き、音声刺激の母音に最も近い日本語の母音を解答用紙の選択肢の中から選び、選んだ日本語の母音としての良し悪しを 7 段階(1=良くない、7=良)で評価した。選択肢は、/p/で始まる子音環境のものは「パ」、「ピ」のよう

にパ行音のカタカナで表示し、/hVt/についてはハ行音のカタカナで表示した。子音環境ごとに分けて、音声は提示し、実験参加者は/pVt/では/p/で始まり/t/で終わる語を聴くが最後の/t/は無視して、その直前までの音を日本語で表記するとどれが一番近いかを選ぶように指示を受けた。各音声刺激は順番を変え 2 回提示された。こうして、6 母音×4 話者×2 回=48 回の試行が各子音環境ごとに作成された。実験参加者が解答を終えたのを確認して、実験者がマウスをクリックして次の試行に移った。ITI は 1,000ms に設定した。

2.3.2　同定実験

実験参加者は、1 回の試行に 1 つの音声刺激を聴き、聞こえた音声を表す選択肢にカーソルを動かし、解答した。選択肢は各子音環境の中で 6 母音で、表 1 に示しているように綴り字を用いた。この中には実験参加者には馴染みのない語や実在しない語が含まれるが、実験参加者には語の意味に左右されずに聞こえたと選択肢を選ぶように指示した。実在しない語については、het は head と同じ母音を持つ語で語尾が d ではなく t になったものであるとか、pul は pulse から s の音を除いたもので、pull とは異なるなどの説明を加えた。各音声刺激は、順番を変え 2 回提示されたこうして、6 母音×4 話者×2 回=48 回の試行が各子音環境ごとに作成された。これで各母音の同定能力は 8 つの試行で測定されることとなった(4 話者×2 回)。ITI は、1,000ms に設定した。また、10 秒間解答しないともう一度同じ音声刺激を聴くことが許された。4 つの子音環境の提示順は参加者ごとに変えられた。

表 1: 同定実験に用いた選択肢一覧

	/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
hVt	heat	hit	het	hat	hot	hut
pvt	Pete	pit	pet	pat	pot	putt
pVn	peen	pin	pen	pan	pon	pun
pVl	peel	pill	pell	pal	pol	pul

2.3.3　弁別実験

AXB 方式を採用した。実験参加者は、1 回の試行に 3 つの音声刺激を聴き、2 番目の音声刺激が最初の音声刺激と最後の音声刺激のどちらと同じカテゴリーの母音を含むか、つまり AAB か ABB かを選択することにより解答した。選択肢は 2 つ用意されていて、AAB の場合は First、ABB の場合は Last を選択するように指示された。6 つの母音のペア、/i/-/ɪ/, /ɛ/-/ɪ/, /æ/-/ɛ/, /æ/-/ɑ/, /æ/-/ʌ/が設定された。それぞれの子音環境で各母音のペアの弁別能力は 12 の試行で測定された。例えば/i/-/ɪ/の場合、/i/-/i/-/ɪ/の試行が 3 つ、/i/-/ɪ/-/ɪ/の試行が 3 つ、/ɪ/-/i/-/i/の試行が 3 つ、/ɪ/-/ɪ/-/i/の試行が 3 つの計 12 の試行で弁別能力が測定された。各試行の 3 つの音声刺激はすべて異なる話者の発話によるもので構成された。ISI、ITI ともに 1,000ms に設定した。同定実験同様 10 秒間解答しないと同じ試行を再度聴くことが許された。4 つの子音環境の提示順は参加者ごとに変えられた。

奇数番号の参加者は、同定実験を先に、偶数番号の参加者は弁別実験を先に行った。

2.4. 結果

2.4.1 知覚同化実験

各子音環境での 6 母音が日本語のどの母音のカテゴリーの音として分類されたかを表 2 に示す。選択率を見ると、常に同じ日本語の母音のカテゴリーに結びつけられた母音はほとんどない。すべての子音環境において最も多く結びつけられた日本語の母音が同じなのは、/ɪ/、/ɛ/だけである。ただし、/pVn/、/pVl/では/ɪ/が「イ」と結びつけられた率は 50%を下回っていて、/hVt/, /pVt/とは明らかに異なることがみてとれる。/æ/、/ɑ/、/ʌ/はいずれの子音環境においても 50%以上の率で同じ日本語の母音とは結びつけられてはいない。図 1 と 2 で見たフォルマント周波数の変動の影響は/pVn/での/æ/や/pVl/での/ɑ/と/ʌ/などに見られる。

表 2: 知覚同化実験の結果:最も多い回答の選択率( )内の数字は日本語の母音として良し悪しの判定の平均

	/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
hVt	ヒ一49.6 (4.8)	ヒ 75.0 (3.9)	へ 60.6 (3.8)	ハ 45.5 (3.3)	ハ 42.4 (3.9)	ハ 39.0 (3.6)
pVt	ピ 53.0 (5.2)	ピ 77.3 (3.8)	ペ 50.4 (4.0)	パ 31.1 (3.8)	パ 41.3 (4.7)	プ 38.6 (3.9)
pVn	ピ 62.5 (4.5)	ピ 40.5 (3.5)	ペ 62.5 (4.2)	ペ 33.7 (3.6)	ポ 24.2 (4.5)	パ 47.4 (4.5)
pVl	ピ 37.1 (3.9)	ピ 39.4 (3.5)	ペ 42.8 (3.1)	ペ 27.7 (3.2)	ポウ 27.7 (4.3)	ポ 34.8 (4.1)

2.4.2 同定実験

6 母音の各子音環境での平均正答率を図 3 に示す。4 つの子音環境と 6 母音を被験者内要因にした反復測定の分散分析の結果、子音環境の主効果[$F(3, 96)=47.46$, $p<.001$]、母音の主効果[$F(5, 160)=40.80$, $p<.001$]、子音環境と母音の交互作用[$F(15, 480=11.38$, $p<.001$]といずれも有意であった。Bonferroni で調整した多重比較の結果、/ɛ/を除くすべての母音が子音環境の影響を受けていることがわかった。/i/と/æ/は共に/pVn/, /pVl/での正答率は/hVt/、/pVt/よりも有意に低かった($p<.001$)。/ɪ/についても同様であったが、更に/pVn/では/pVl/よりも有意に正答率が低かった($p=.024$)。/ɑ/は、逆に/hVt/での方が/pVt/、/pVn/よりも有意に正答率が低かった(それぞれ $p=.032$、$p=.039$)。また、/ʌ/は、/pVt/の方が他の 3 つの子音環境に比べて有意に正答率が高かった

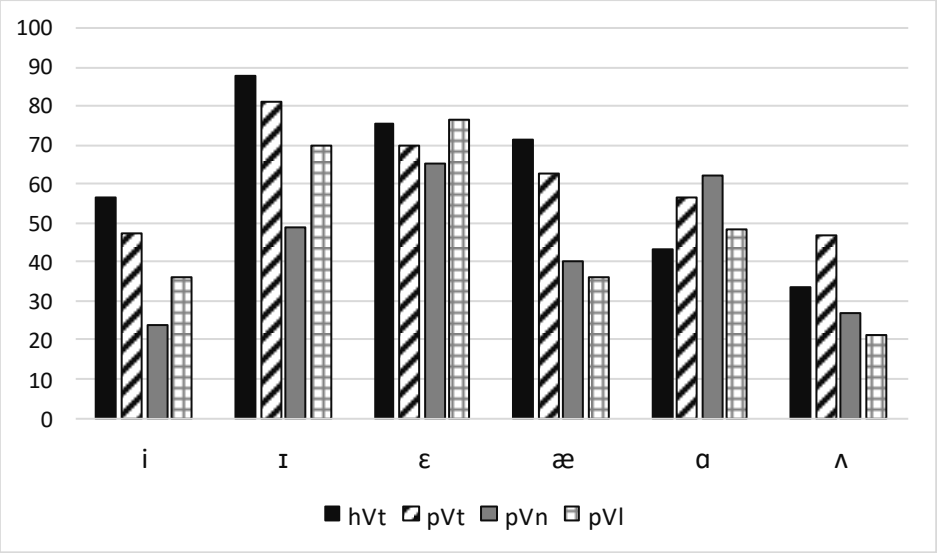


図 3: 同定実験の正答率(%)

2.4.2 弁別実験

6 母音の各子音環境での平均正答率を図 4 に示す。

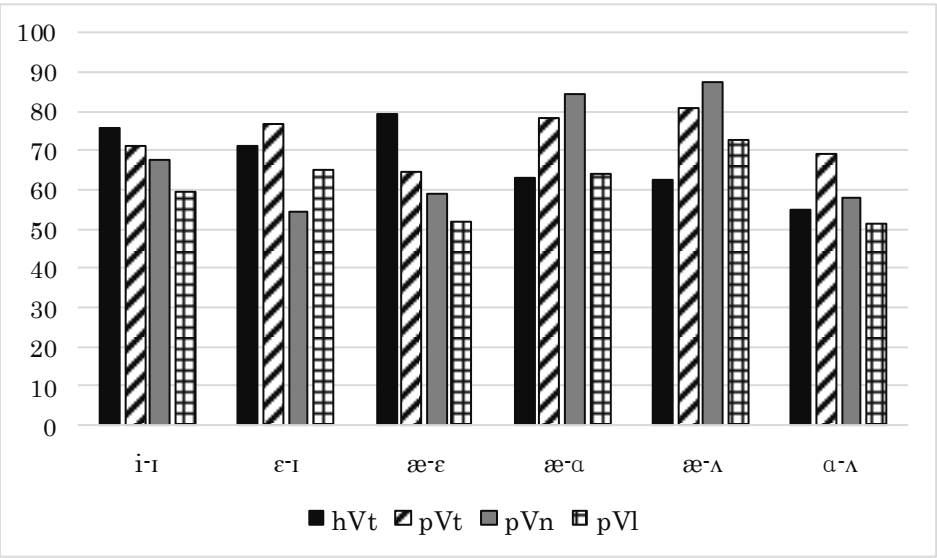


図 4: 弁別実験の正答率(%)

同定実験同様、4 つの子音環境と 6 つの母音のペアを被験者内要因にした反復測定の分散分析を行った。その結果、子音環境の主効果[$F(3, 96)=13.46$, $p<.001$]、母音のペアの主効果[$F(5, 160)=21.63$, $p<.001$]、子音環境母音のペアの交互作用[$F(15, 480)=12.96$, $p<.001$]のいずれも有意であった。全体を通して、/pVt/での正答率が他の 3 つの子音環境に比べて $p<.05$ の水準で高く、逆に/pVl/での正答率は他の 3 つの子音環境よりも $p<.05$ の水準で低かった。Bonferroni で調整した多重比較の結果、以下に示すように全ての母音のペアの正答率に少なくても

p<.05 の水準で有意差があった。

表 3: 弁別実験の多重比較の結果

/i/-/ɪ/	/ɛ/-/ɪ/	/æ/-/ɛ/	/æ/-/ɑ/	/æ/-/ʌ/	/ɑ/-/ʌ/
pVl<hVt, pVt	pVt>pVn, pVl hVt>pVn	hVt>他全 pVt >pVl	hVt<pVt, pVn pVl<pVt, pVn	hVt<pVt, pVn pVl<pVn	pVt>他全

3. 結論

以上見てきたように, 母音に後続する/n/と/l/は日本語母語話者の英語の母音の知覚に大きく影響することがわかった。3 つの実験結果がどのように関連しているかは、表や図を見ただけではわかり辛いが、例えば/æ/は、冒頭で述べたように鼻音の前では前寄り・高めの位置に移動する。この影響が表 2 にも表れている。/æ/は一般に「ア」に近い母音とされているが、鼻音の前では「エ」に近い母音に聞こえるため、同定実験での正答率は低くなる。/æ/が/ɛ/に近づくため、弁別実験での/æ/-/ɛ/の正答率は下がり、逆に/ɑ/や/ʌ/とは離れるため、/æ/-/ɑ/と/æ/-/ʌ/の正答率は高くなる。音節末の/l/には音響的には母音との分節化が難しいが、実験参加者がどこまでを母音として、どこからを/l/として知覚したのかという問題が残る。

4 人の話者の発話を音声刺激として使用したが、4 人の音声刺激に対して同じ反応を示した訳ではない。知覚の実態を解明するには個々の音声刺激に対する回答を分析する必要がある。今回は日本語の母語話者であること以外に特に条件を設けずに参加者を募り、実験結果に大きな個人差が生じた。個々の実験参加者ごとに実験結果の関連を精査する必要がある。本研究は、科学研究費補助金 C（16K02650）の補助を受けたものである。

参考文献

Beddor, P. S. (1993) The Perception of Nasal Vowels, In Huffman, M.K. & Krakow, R. A. (eds.) *Nasals, Nasalization, and the Velum* (*Phonetics and Phonology* Volume 5) pp. 171-196. .Academic Press.

Krakow, R A., Beddor, P. S., Goldstein, L. M. & Fowler, C. A. (1988) Coarticulatory Influence on the Perceived Height of Nasal Vowels, *Journal of the Acoustical Society of America* 83, 1146-1158.

Labov, W., Ash, S. & Boberg, C. (2005) *Atlas of North American English: Phonetics, Phonology & Sound Change.* Mouton de Gruyter.

Nozawa, T. & Wayland, R. (2012) Effects of Consonantal Contexts on the Discrimination and Identification of American English Vowels by Native Speakers of Japanese, *Journal of the Japan Society of Speech Sciences* 13, 19-39.

Olive, J. P., Greenwood, A., & Coleman, J. (1993) *Acoustics of American English Speech.* New York: Springer-Verlag.

P05

日本語母語話者による中国語声調の知覚と生成

楊 姝怡（神戸大学大学院）・山田 玲子（ATR/神戸大学）
syuiyang@gmail.com

1. はじめに

中国語には四つの声調の区別があるが(Chao, 1948)，非母語話者は知覚と生成の両面で声調を混同する場合がある(Chuang *et al*., 1972; Kiriloff, 1969; Wang *et al*., 1999; Wang *et al*., 2003; Yang & Akahane-Yamada, 2016; 他). また, その混同の割合は声調によって異なる. Wang *et al*. (1999) はアメリカ英語母語話者を対象として，四つの声調のうちの 2 つを組み合わせたミニマル・ペアの単音節語（例えば，běi, bèi）を刺激語とした二肢強制選択課題の知覚実験を行った．その結果，正答率は一声，三声＞四声＞二声の順に低下した．また，Wang *et al*. (2003)ではアメリカ英語母語話者に単音節語を発音させ，その音声を中国語母語話者に一声～四声，または「どれでもない」の５つの選択肢から判定させた．その結果，発話の発音明瞭度（話者が発音した声調と判定者が選んだ声調が一致した率）は一声＞四声＞二声＞三声の順に低下した.

本研究では日本語母語話者を対象として，中国語声調の知覚実験と発音の収録を行い，声調による難易度およびその知覚と生成の関係について検討した.

2. 方法

2.1. 実験参加者

日本語母語話者 30 人（男性 14 人，女性 16 人；18~28 歳，平均 21 歳）が参加した．全員中国語学習経験はなかった．また，中国語を使う地域での一か月以上の居住経験もなかった.

2.2. 刺激音

4 つの声調で対立する中国語単音節語（例えば，bā, bá, bǎ, bà）を 1 組とし，36 組 144 語を刺激語とした．刺激語を 1 人の中国語母語話者（女性，23）が読み上げ，サンプリング周波数 44.1kHz，量子化精度 16bit でデジタル化し，音声ファイルとして格納したものを音声刺激とした.

2.3. 手続き

実験開始時に，中国語－日本語のバイリンガルである実験者が中国語声調の説明が日本語で記されたプリントを見せ, 説明を読み上げるとともに 4 つの声調の音声を聞かせつつ，声調の違いを教示した．この教示の後，実験を開始したが，実験では刺激の提示，反応の取得はすべてパソコンで制御した.

知覚実験では，実験参加者は静かな部屋でパソコンに向かってヘッドホンを付けて座り，キーボードを使って回答した．各トライアルでは 1 つの刺激語に対してそれと声調で対立する 3 語をディストラクタとし, 4 つの選択肢がピンインでパソコンのディスプレイ上に表示された（図１）．選択肢はトライアル毎にランダムな順に配置された．実験参加者には聞

こえた音声と一致した選択肢を選ぶように教示した．刺激音はトライアルの冒頭で 1 回だけ再生され，繰り返し聞くことはできなかった．144 の刺激音をランダムな順で呈示した．

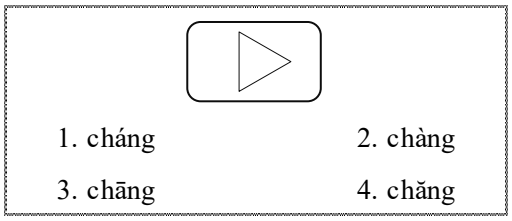


図 1: 知覚実験の画面例

音声の収録においては実験参加者は刺激音を聴取し，それをまねして発音した．実験参加者には声調の変化に注意して発音するように，あらかじめ教示した．その発音をサンプリング周波数 44.1kHz，量子化精度 16bit でデジタル化し，音声ファイルとして格納した．後に 6 人の中国語母語話者（女性 3 人，男性 3 人；22~33 歳，平均 25 歳）がその音声を声調で対立する 4 種類のピンインのうちのどれに聴こえるか，前項の知覚実験と同じ強制選択課題を用いて評定した．一人の評定者は 4,320 全音声ファイル（144 刺激語×30 人）を話者（実験参加者）ごとのブロックに分けて評定した．

3. 結果

知覚実験について，参加者ごと声調ごとの平均正答率を求めた．声調（一声～四声）を要因とし，平均正答率を従属変数として F 検定を行った結果，声調によって正答率は有意に変動した[$F\ (3,116) = 12.681$, $p<0.0001$]（図 2 左）．一声～四声の正答率はそれぞれ 74.6%，58.4%，96.8%，67.9%であった．

生成では，各発話音声について 6 人の評定者のうち刺激語と同じ語を選択した割合を求めた．本研究ではこの割合を発音明瞭度と呼ぶ．声調（一声～四声）を要因とし，発音明瞭度を従属変数として F 検定を行った結果，声調（一声～四声）の効果は有意だった[$F\ (3,116) = 29.270$, $p<0.0001$]（図 2 右）．一声～四声の正答率はそれぞれ 96.3%，87%，78%，98.8%であった．

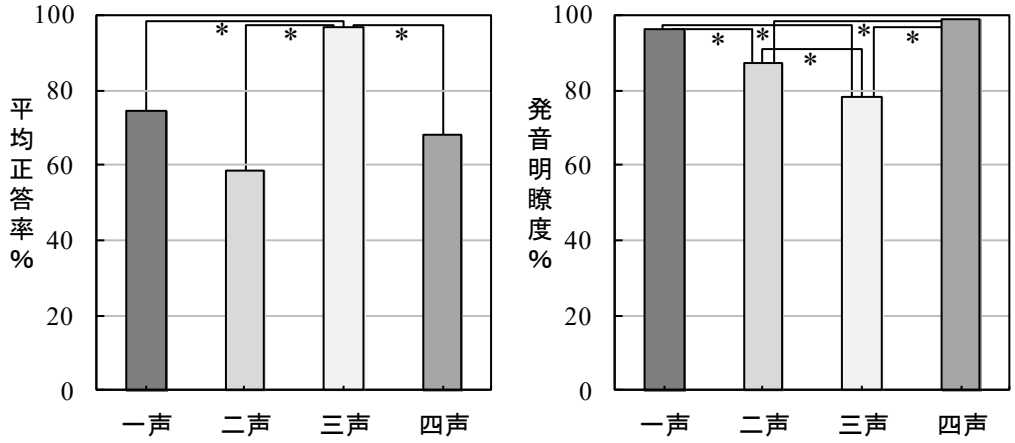


図 2: 知覚実験(左)と生成(右)の結果 *p<0.001

4. 考察

知覚実験の結果では，平均正答率は三声が一声，二声，四声より有意に高かった(三声＞一声, 二声, 四声)．生成の明瞭度は三声が一声，二声，四声より有意に低く，四声，一声＞二声＞三声の順に低下した．知覚と生成の結果を比較すると，三声は知覚の正答率は最も高いが発音明瞭度は最も低く，一声と四声は知覚の正答率は低いが発音の明瞭度は 100%に近かった．知覚は容易だが発音が困難，またはその逆の音が存在することが示唆されたといえる．

また，知覚の結果は Wang *et al*. (1999)の知覚実験の結果とは，全体的な傾向は類似していたが，本研究では三声の正答率が顕著に高かった（図 3 左）．生成の結果は Wang *et al*. (2003)のアメリカ英語母語話者の生成の難易度と全体的な傾向は類似していたが，本研究では声調間の差が小さかった．

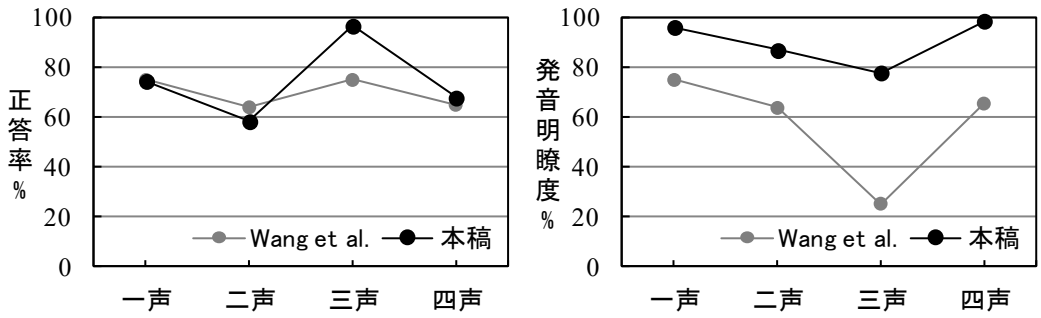


図 3: Wang 等によるアメリカ英語母語話者の結果と日本語母語話者を対象とした本研究結果の比較．左が知覚の比較（Wang *et al*., 1999），右が生成の比較（Wang *et al*., 2003）．

Wang 等と本研究と比較すると，課題が多少異なるので直接の比較はできないものの，アメリカ母語話者と日本語母語話者で声調の効果について生成, 知覚とも類似した傾向を示したことは, 声調の難易度は両母語話者に共通した傾向である可能性を示した．また，本研究では中国語を全く知らない日本語母語話者を対象としたが, 特に生成について高スコアを示したことは興味深い．声調のみならず中国語音声に関する知識を持たないは, ピッチの変動を真似することで一時的に明瞭度が上昇している可能性がある．母語である日本語がピッチアクセントであることも関係しているかもしれない．今後，中国語学習過程との関連を検討することにより, 効果的な発音習得方法について検討したい.

参考文献

Chao, Y. R. (1948). Mandarin primer: An intensive course in spoken Chinese. Harvard University Press.

Chuang, C. K., & Hiki, S. (1972) "Acoustical features and perceptual cues of the four tones of standard colloquial Chinese", *The Journal of the Acoustical Society of America* 52(1A), 146-146.

Kiriloff, C. (1969) "On the auditory perception of tones in Mandarin", *Phonetica* 20(2-4), 63-67.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999) "Training American listeners to perceive Mandarin tones", *The Journal of the Acoustical Society of America* 106(6), 3649-3658.

Wang, Y., Jongman, A., & Sereno, J. A. (2003) "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training", *The Journal of the Acoustical Society of America* 113(2), 1033-1043.

Yang, S., & Akahane-Yamada, R. (2016) "Effects of perception and vocabulary training of Mandarin tones for native speakers of Japanese: Pre-, post-, and retention test comparison", *The Journal of the Acoustical Society of America* 140(4), 3392-3392.

P06

自然談話における終助詞「ね」と「よ」のイントネーション

張海霞(東京学芸大学大学院)　　白勢彩子(東京学芸大学)
m161419k@st.u-gakugei.ac.jp　　shirose@u-gakugei.ac.jp

1. はじめに

日本語の会話において、終助詞の「ね」と「よ」は高頻度に出現し、話者の意図を伝達する重要な役割を担っている。イントネーションにより、「ね」と「よ」は意味機能に変化があるが、二つの関係を体系的に議論した研究は多くない。そこでイントネーションと意味機能の関係をより詳しく記述するため、我々は文に上昇調、平坦調、下降調三パタンの文末イントネーションを付与し、意味判定実験を実施した(Zhang and Shirose,2016)。その結果、「ね」は下降調を伴うと「不自然」と高く判断され、「よ」は下降調を伴うと、「困惑」「非難」のような不快な意味に判断される傾向があることが分った。

本稿では、前回の聴取実験を踏まえ、自然談話における終助詞「ね」と「よ」及びイントネーションパタンの出現状況を調べ、意味機能の関係を検討する。

2. 先行研究

終助詞「ね」と「よ」の意味機能に関しては、従来多くの研究が行われてきた。金水(1993)「ね」は話し手は聞き手と知識を共有するため、「確認」「同意を求める」の意味機能を持ち、「よ」は聞き手の知らない知識を教えるために用いられると述べた。メイナード(1992)は、話し手と聞き手のそれぞれの情報量の状況から、終助詞「ね」と「よ」の使用条件を呈示した。金水とメイナードのように、情報帰属あるいは情報量の視点から、終助詞「ね」「よ」の意味機能を議論した研究が多いが、イントネーションとの関係を論じた研究は多くない。終助詞の意味機能は終助詞の元々の意味とイントネーションとの相互が関わると述べた片桐(1997)、自己内省で終助詞「ね」と「よ」の意味を分析した小山(1997)音声合成で「ね」の機能の境界を論じた杉藤（2001)、ピッチ曲線でイントネーションを分類し、意味機能を分析したデイヴィッド（2013）がある。しかし、自然談話において、終助詞「ね」と「よ」のイントネーションがどのように現れるについての調査はほぼない。

本稿は自然談話を前提として、終助詞「ね」「よ」のイントネーション様子を呈示しつつ、意味機能との関係も検討する。

3. 方法

3.1 資料

『千葉大学 3 人会話コーパス』は日常場面に近い自然な会話を収めた、大学生・院生・ポスドクを含む 3 人からなる友人同士 12 組の雑談、7599 発話のデータである。今回の調査ではこのコーパスを利用し、終助詞「ね」と「よ」の生起頻度を調査した。

3.2 対象となる文の選別基準

『日本語国語大辞典』で、「終助詞は文の終わりにあって、命令・疑問・反語・願望・禁止などの意味を決定し、または陳述の意味を強めたり、感動を表わしたりする助詞」と定義していることから、文の中間で使用される文を対象外とする。今回の調査でも、複合助詞、接続助詞との接続、疑問詞、相槌表現も対象外とした。具体的な標準は表 1 に示す。対象となる文は表 2 のような文である。

表 1　対象外とする基準

基準	例
A　文間で使用される「ね」「よ」	私はね
B　複合助詞	よね、かね
C　単独使用	ね
D　接続詞と接続する場合	思うけどね
E　疑問詞	いつ…ね
F　相槌表現	そうだね

表 2　対象となる文

番号	例
1	水冷売ってるよ
2	今週末ぐらいだったね
3	大昔だね
4	怖かったよ
5	すごいね

3.3 文末イントネーションの分類基準

音響分析ソフトウェア Praat に基づくピッチ曲線の視察および聴覚印象により、終助詞「ね」「よ」のイントネーションパタンを分類した。

今回の調査では、終助詞「ね」と「よ」部分のイントネーションを上昇調、平坦調、下降調、複合音調に分類した。上昇調と平坦調の区別について、10Hz で分けた。「ね」と「よ」内の音調は 10Hz 以上上昇した場合、上昇調をし、10Hz 以内で変動する場合、平坦調とした。複合音調も 10Hz を基準にし、2 か所以上で 10Hz 以上変動すると、複合音調とした。ピッチ曲線を抽出されない場合もあり、知覚による判断とした。以下は上昇調(図 1)、平坦調(図 2)、下降調(図 3)、複合音調(図 4)のサンプル(Praat 分析画面)である。

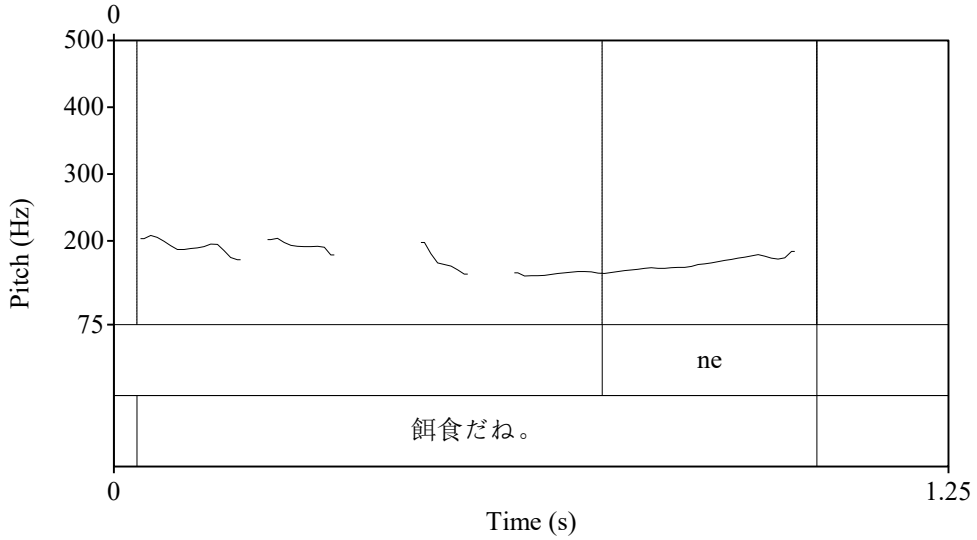


図 1　上昇調

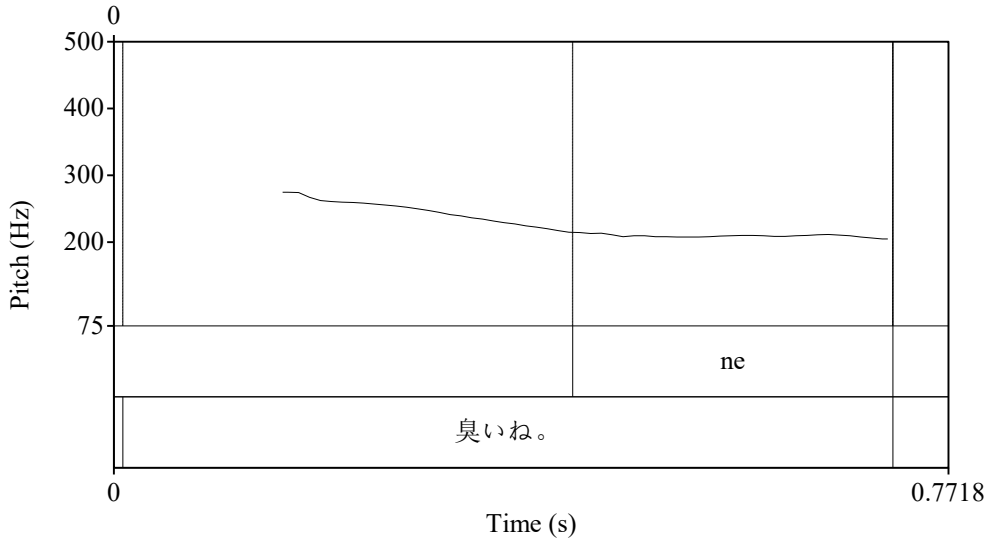


図 2　平坦調

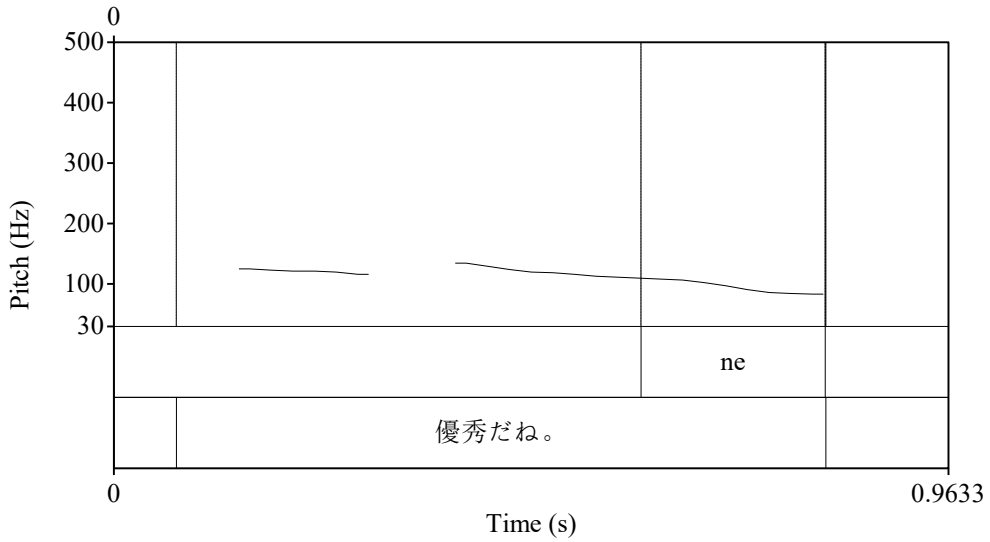


図 3　下降調

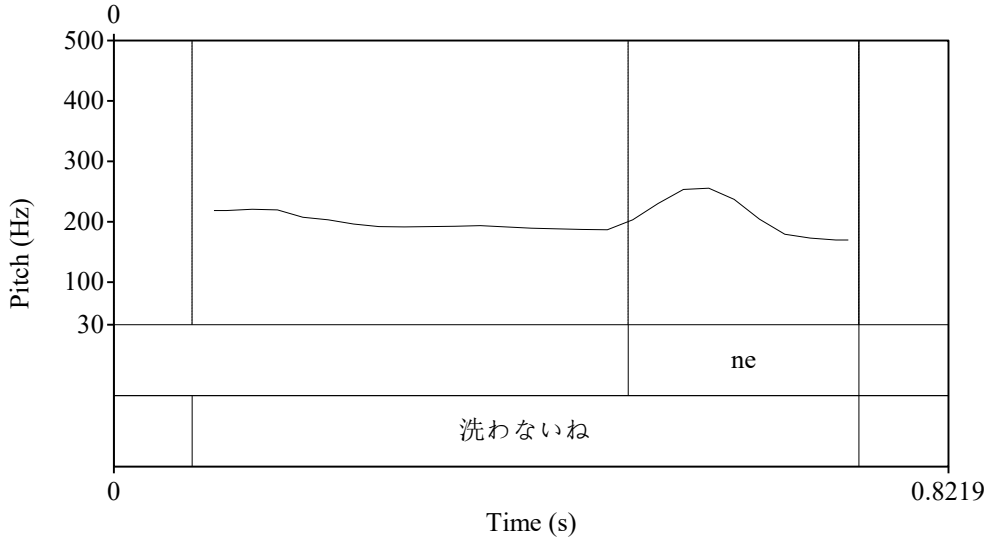


図 4　複合音調

4　結果と考察

3.2 の選考基準により、「ね」と「よ」、それぞれ 84 発話と 145 発話が抽出された。終助詞「よ」の使用数は「ね」より約 1.7 倍多く生起していることが分かった。イントネーションの分類の結果、終助詞「ね」と「よ」の出現数をそれぞれ図 5 と図 6 に示す。

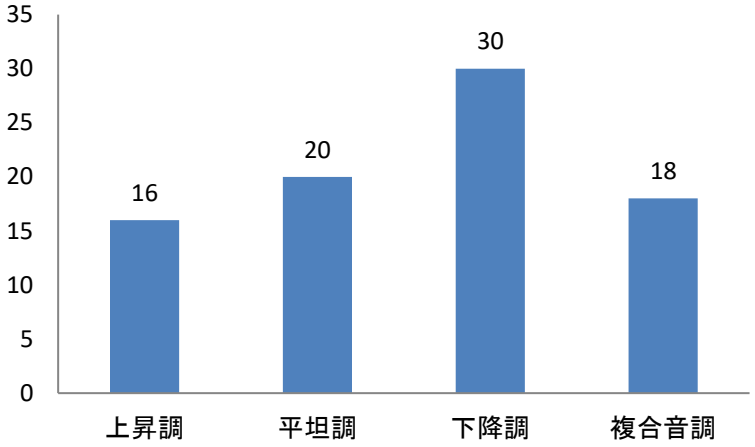


図 5　終助詞「ね」のイントネーションパタンの出現数

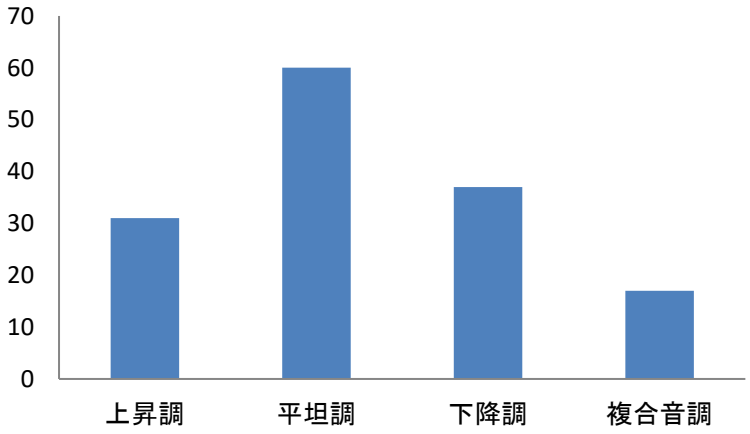


図 6 終助詞「よ」のイントネーションパタンの出現数

終助詞「ね」の出現状況について、上昇調 16(19.1%)、平坦調 20(23.8%)、下降調 30(35.7%)、複合音調 18(21.4%)という結果となった。下降調は他の音調より多く出現することが分った。前回の聴取実験では、下降調になると、不自然(61.5%)と判断されやすく、イントネーションによる意味機能の変動が大きくないという結果があった。ただし前回の聴取実験では、下降調になると、「感心」というような意味が高く判断される傾向があった。今回の調査で対象とした自然談話では、「感心」のような自己感情を表す意図が多く出現し、このことが下降調の生起に反映したと考えられるだろう。

終助詞「よ」の出現状況については、上昇調 31(21.4%)、平坦調 60(41.4%)、下降調 37(25.5%)、複合音調 17(11.7%)という結果となった。複合音調はあまり出らず、平坦調が主に出現したことが分った。前回の聴取実験では、「よ」の上昇調、平坦調、下降調について、不自然と判断されにくく、どれでも許容して使用いられるという結果があった。下降調になると「困惑」「非難」のような不快な意味になる傾向があり、「よ」の意味機能はイントネーションにより変動しやすいとの結果があった。今回の調査結果から見ると、平坦調を主に出現することで、話し手と聞き手が談話をスムーズに進められるように、意味機能を抑え、イントネーションを採用したと考えられる。

5　まとめ

コーパス分析の結果から、自然談話において、終助詞の「よ」は「ね」より多く使用されていることが分った。終助詞「ね」について、非上昇調が主に出現し、「感心」というような自己感情を表す発話が多かったと考えられた。「よ」については、平坦調が主に出現することは、自然談話を順調に続けるため、意味機能を抑えたと考えられるだろう。自然談話における、終助詞の意味機能とイントネーションの相互関係の詳細については、今後の課題としたい。

参考文献

金水　敏（1993）「終助詞ヨ・ネ」『言語』第 22 巻第 4 号, pp.118－121.

メイナード・泉子・k（1993）『会話分析』pp.101－109,くろしお出版.

伊豆原　英子（1993）「「ね」と「よ」再考－「ね」と「よ」のコミュニケーション機能の考察から」　『日本語教育』80 号, pp.103－114.

伊豆原　英子（2001）「「ね」と「よ」再再考」『愛知学院大学教養部紀要』第 49 巻第 1 号, pp.35－49.

小山　哲春（1997）「文末詞とイントネーション」『文法と音声』pp.97-119,くろしお出版.

片桐　恭弘（1997）「終助詞とイントネーション」『文法と音声』pp.235-256 , くろしお出版

杉藤　美代子（2001）「終助詞「ね」の意味・機能とイントネーション」『文法と音声Ⅱ』pp.3-16,くろしお出版.

大島　デイヴィッド　義和（2013）「日本語におけるイントネーション型と終助詞機能の相関について」『国際開発研究フォーラム』43, pp.47-63

Zhang and Shirose (2016) A study of Japanese modal particles “ne” and “yo”,Journal of Acoustical Society of America, 140-4,P3394.

P07

F0, long-term formants and LTAS in Korean-English Bilinguals

Sylvia Cho, Murray J. Munro (Simon Fraser University)
sylvia_cho@sfu.ca, mjmunro@sfu.ca

1. Introduction

Voice quality measures such as speaking fundamental frequency (SFF), long-term formants (LTF), and long-term average speech spectra (LTAS) have been useful for speaker identification of monolingual speech samples (Nolan & Grigoras, 2005). However, the existing studies on SFF, LTFs, and LTAS either lack bilingual data or have reported obscure or confusing results. It is thus largely unknown whether the acoustic properties of speech samples from a bilingual speaker differ according to the language being spoken. Such comparisons have important implications for the practice of forensic speaker identification, as bilinguals may or may not behave like their monolingual counterparts in their productions. The current study therefore aims to contribute to this line of work by (1) extending SFF comparisons to Korean-English bilinguals, (2) conducting a comparison of bilingual LTFs that employs LTF histograms (as well as mean values), and (3) confirming whether LTAS differences between monolingual speakers of Korean and English are also found in Korean-English bilinguals.

2. Production Experiment

2.1. Methods

2.1.1 Participants

Ten Korean-English bilingual speakers (5M 5F) were recruited at Simon Fraser University. All subjects spoke standard Seoul Korean and Western Canadian English. Their age ranged from 18 to 36 (Mean age: 25), and their age of arrival (AOA) in Canada ranged from 0 to 12 years (Mean AOA: 5.8). Each subject was screened prior to the recording to ensure that they met the following criteria: (1) they were judged to be proficient Korean-English bilinguals by one of the authors who is proficient in both languages (2) they did not report any hearing or speech impairments; (3) they had no signs of sickness from colds or other respiratory problems on the day of the recording.

2.1.2 Procedures

A total of four speech samples were taken from each subject. The subjects were instructed to read a passage out loud (i.e., The North Wind and the Sun), at a comfortable pace in both Korean and English. They were also asked to produce extemporaneous narratives in both languages, in which they were instructed to describe a picture story (Derwing, Munro, Thomson, & Rossiter, 2009).

3. Results and Discussion

3.1. Speaking Fundamental Frequency (SFF)

Speaking fundamental frequency (SFF) was measured using the autocorrelation method in Praat (Boersma & Weenik, 2016). Paired-sample *t*-tests[1] were then carried out to assess the existence of a language effect on SFF. These results are shown in Figures 1 and 2.

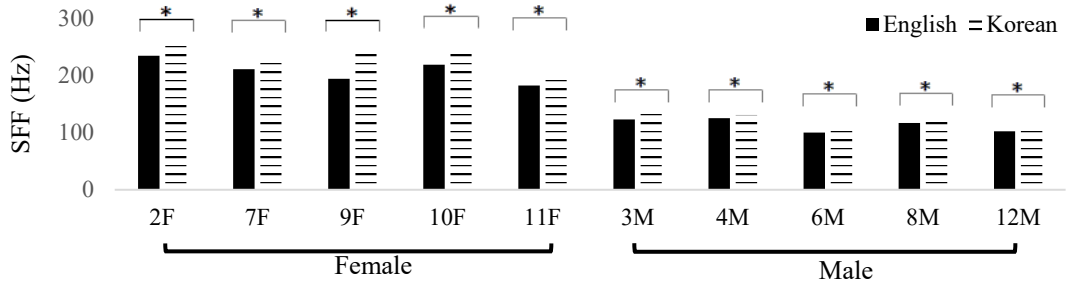


Figure 1: Average SFF (in Hz) for passage reading in Korean and English

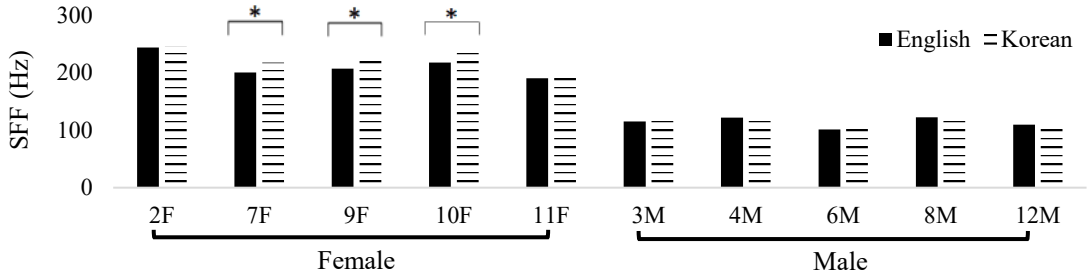


Figure 2: Average SFF (in Hz) for extemporaneous speech in Korean and English (* $p < 0.05$)

As displayed in Figures 1 and 2, significant SFF differences were found between English and Korean in the passage reading, but this language effect was not as robust in the extemporaneous data (i.e., only 3 out of 10 speakers adjust their SFF significantly). With the passage reading, there was a mean F0 difference of 15 to 45 Hz between the languages for the female subjects, and for the male subjects, there was a difference of 5 to 20 Hz. In the extemporaneous speech task, there was a maximum of 15 Hz difference in the females, and up to a 5 Hz difference in the males. This indicates an effect of speaking style on SFF, and that casual or spontaneous speech is produced with smaller between-language differences. In terms of the direction of the SFF change, there was either no statistical difference between the bilinguals' production of Korean and English, or English was produced with a lower SFF. Regarding this observed pattern, it is possible that subjects produced English with a lower

[1] Each speech sample was parsed into 20 stretches of equal duration. The mean F0 value for each stretch was obtained, and these values (i.e., 20 data points for Korean; 20 data points for English) were used in the paired-sample *t*-tests.

SFF due to a language-dependent tendency. That is, English has been associated with a relatively low SFF, compared to that of other languages (Hanley & Snidecor, 1967). This tendency seems to be more readily seen in read speech, as opposed to extemporaneous speech, which better approximates natural speaking conditions.

3.2. Long-term Formants (LTF)

LTFs were computed for speech samples of equal duration (i.e., first 30 seconds of speech sample) using formant frequencies chosen by the automatic formant tracker in Praat, which uses linear predictive coding. In terms of mean formant values, there were very small within-subject differences. This study finds that the shapes of the formant frequency distributions are more indicative of speaker identity than are the means. For example, the histograms in Figures 3 and 4 display the distributions of formant value ranges (displayed in 100 Hz bins) for two speakers that have very similar mean formant values (i.e., F1: 400 Hz, F2: 16-1800Hz, F3: 2800Hz). While their average formant frequencies are similar, Subject 4M, has a more prominent F1 peak, and his F3 peak is skewed more to the left.

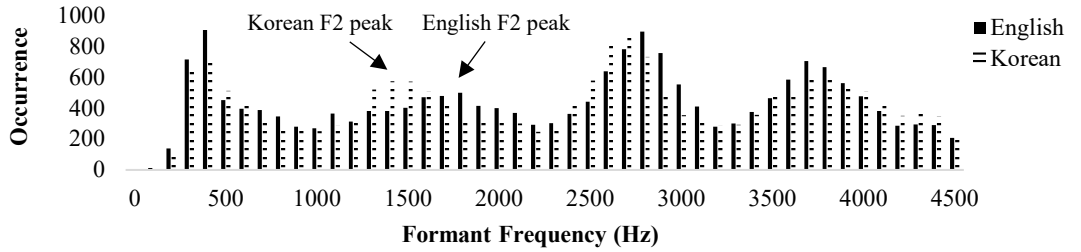


Figure 3: Read Speech LTF for Subject 3M

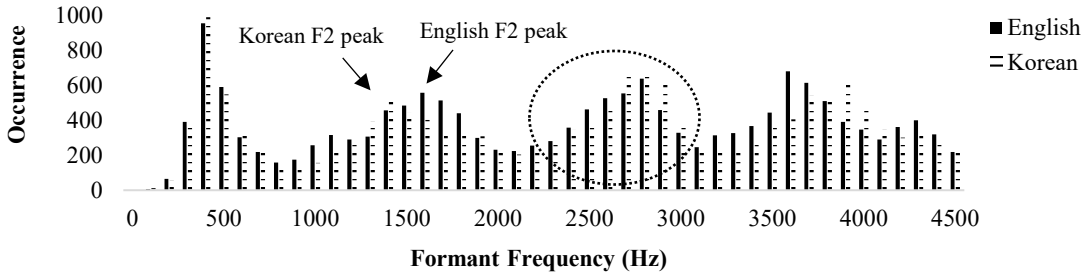


Figure 4: Read Speech LTF for Subject 4M

As shown in the histograms above, there is also a small effect of language across bilingual LTFs. That is, Korean productions are often associated with a lower F2 peak. This is consistent with acoustic studies that report slightly lower F2 values for Korean vowels in comparison to English vowels (Yang,

1996). A strong effect of speech style was not found with bilingual LTFs. Altogether, the current study finds that LTFs are mostly speaker-specific with small language-specific effects.

3.3. Long-term Average Speech Spectra (LTAS)

There appear to be no existing studies on bilingual LTAS, but a study that compared LTAS of twelve languages[2] found that most of the variations among languages are smaller than 3 dB in magnitude and occur within the range of 200 to 6300 Hz (Bryne et al., 1994). It was thus suggested that a "universal" LTAS can be applied across languages. A more recent study (Noh & Lee, 2012) compared Korean LTAS to that of English, and found that the LTAS of Korean speakers show significantly lower intensity levels in frequencies above 2000 Hz, except at 4000 Hz. The lower intensity levels found in Korean are thought to stem from lower occurrence of phonemes with high frequency components (i.e., fewer fricatives in Korean). This suggests that for languages such as Korean, the universal LTAS is rendered inappropriate. The current study examines whether the trends found for monolingual English and Korean speakers also hold true in bilingual speakers.

LTAS were computed using Praat to observe overall spectral properties of the continuous speech samples. Each speaker's samples were then normalized for peak intensity (dB) to permit straightforward comparisons. This study finds that most speech spectra show maximum intensity around 500 Hz, and a decrease is seen in the higher frequencies. Similar to the monolinguals in Noh and Lee's study, many bilingual speakers (6 out of 10) showed significantly lower intensity levels in frequencies above 2000 Hz (except at 4000 Hz) in their production of Korean (see Figure 5). There was a 3 to 5 dB difference from 2000 to 4000 Hz, and a difference of 5 to 10 dB in the higher frequencies.

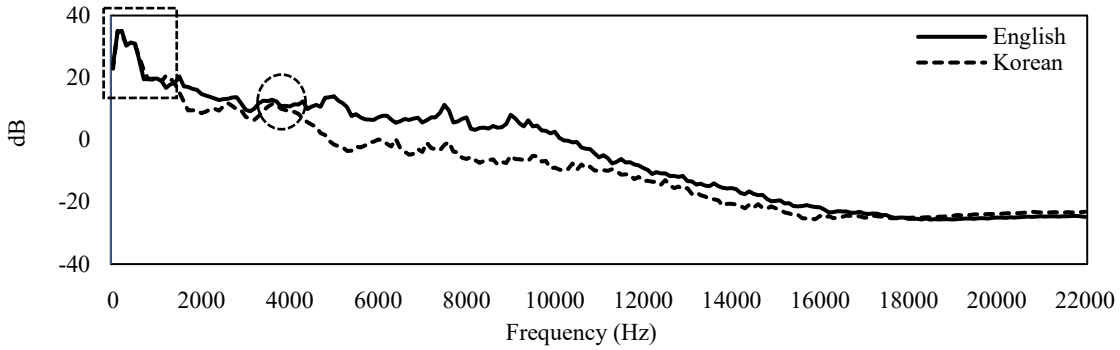


Figure 5: Extemporaneous Speech LTAS for Subject 4M

It should be noted that this pattern was not observed across all speakers. Some bilinguals (4 out of 10) had very similar intensity levels in the two languages up to 4000 Hz (see Figure 6). This suggests that

[2] English (several dialects), Swedish, Danish, German, French, Japanese, Cantonese, Mandarin, Russian, Welsh, Singhalese, Vietnamese

there are some within-speaker characteristics that influence bilingual LTAS – which is different from what is found across monolingual speakers (i.e., between speakers). The reasons for such a distribution are not clear at the moment, but it seems that the bilingual speakers that were impressionistically more dominant in one language (i.e., better at either Korean or English) demonstrated such patterns.

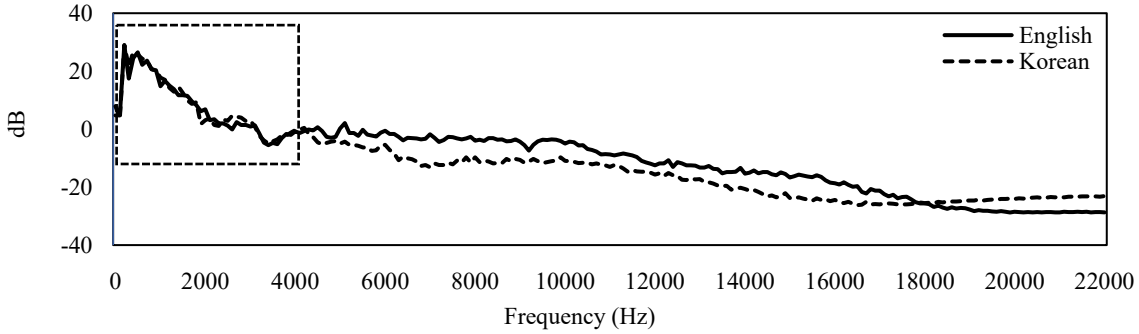


Figure 6: Extemporaneous Speech LTAS for Subject 2F

Furthermore, this study found nearly identical LTAS for bilingual speakers that had clearly lower proficiency in one of the languages. Figure 7 shows the LTAS of a speaker that was omitted from the scope of this study due to his lack of native-like proficiency in English (late-onset bilingual, AOA: 18 years old).

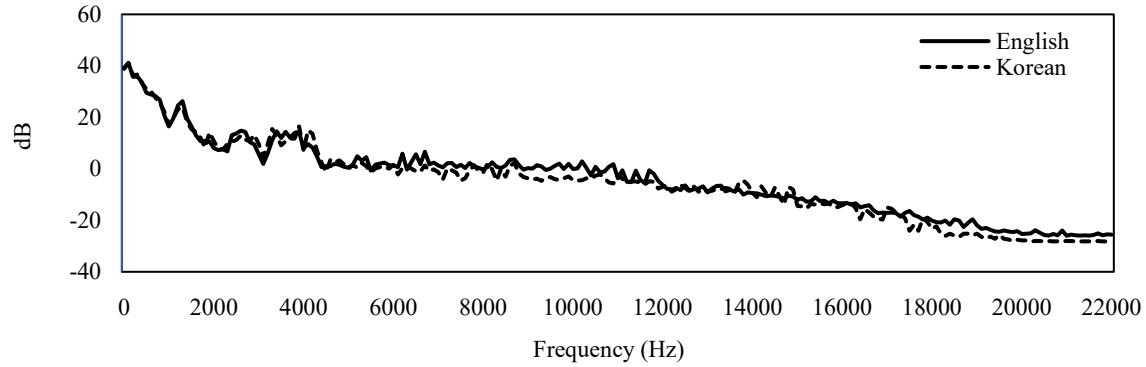


Figure 7: Extemporaneous Speech LTAS for Subject 5M

As can be observed from the figure above, energy differences seen across other Korean-English bilinguals were not found for Subject 5M, namely a subject who is not very proficient in one of the languages. It is plausible that such speakers produce the less-proficient language in much the same manner as their native language. Such findings suggest that LTAS analysis can be somewhat useful in speaker identification cases involving multilingual samples, given that the speaker lacks proficiency in one of the languages spoken. Lastly, there was little indication of an effect of speech style on LTAS distributions.

4. Conclusion

This study investigated the usefulness of SFF, LTF, and LTAS analyses in bilingual speaker identification. Among the different acoustic measurements, formant frequency estimates across long speech samples were found to be most reliable and speaker-specific. Speaking style was also found to be important; stronger SFF correlations were found between languages in extemporaneous speech. Such results suggest that multilingual samples may serve some useful purpose in speaker identification cases that involve bilingual speakers. Future research could further expand on the results obtained from this project by applying the methods to other bilingual populations, such as Japanese-English bilinguals. One previous study finds that SFF is consistently higher in Japanese for Japanese-English bilinguals (Graham, 2015), and it might be the case that, like Korean-English bilinguals, there is some effect of language in the bilingual LTFs given the small vowel inventory of Japanese.

References

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.23) [Computer program]. Retrieved from http://www.praat.org/

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox. R., ... Ludvigsen, R. (1994). An international comparison of long-term average speech spectra. T*he Journal of the Acoustical Society of America, 96*(4), 2108-2120.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31(4)*, 533-557.

Graham, C. (2015). Fundamental Frequency Range in Japanese and English: The Case of Simultaneous Bilinguals. *Phonetica, 71*(4), 271-295.

Hanley, T. D., & Snidecor, J. C. (1967). Some acoustic similarities among languages. *Phonetica, 17*, 141-148.

Heeren, W., Van der Vloed, D., & Vermeulen, J. (2014). Exploring long-term formants in bilingual speakers. *Proceedings from the 2014 Conference of the International Association for Forensic Phonetics and Acoustics*. Zurich, Switzerland.

Krebs, P., & Braun, A. (2015). Long Term Formant measurements in bilingual speakers. *Proceedings from the 2015 Conference of the International Association for Forensic Phonetics and Acoustics*. Leiden, Netherlands.

Noh, H., & Lee, D. (2012). Cross-Language Identification of Long-Term Average Speech Spectra in Korean and English. *Ear and Hearing, 33*(3), 441-443.

Nolan, F., & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law, 12*(2), 143-173.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24*(2), 245-261.

P08

東京出身者の母音フォルマントの世代間比較

脇田 佳幸（NIEMAN INTERNATIONAL SCHOOL）
wakida.13y@gmail.com

1. はじめに

本研究は『日本語話し言葉コーパス』（CSJ）中の東京出身者の母音を音響分析することで、母音のフォルマントに世代差がある可能性を指摘する。日本語（共通語）の母音の発音に世代差がある可能性については、いくつか短い指摘があるものの、それを音響分析で検証した研究はほとんどない。そこで本研究は、母音フォルマント（F1・F2）に世代差が見られるかどうかを検証する。

2. 母音の世代差に関する指摘

日本語（共通語）の母音に音変化が見られる可能性について、音声の専門家が複数指摘を行っている。例えば、ウの円唇性の度合いの変化について、杉藤美代子（1997: 3-4）は、東京の若い世代に非円唇ではなく円唇のウの発音があるとする。また、アの変化について、井上史雄（1989: 111）は、「現代日本語では母音 a が前寄りになり、おそらくその影響で、e が狭くなる傾向がみられる」と指摘する。さらに、柴田実（2005: 66）は、若い世代のアはあごの開きが不足して明るさがない、と主張している。

いずれも音声に関わる専門家が母音の変化を指摘したものである。しかし、これらを受けて、母音の世代差を実証的に検証した研究はほとんどない。そこで本研究は、日本語の母音を音響分析することで、母音に世代差があるのかどうかを実証的に示すことを目指す。

3. 使用データ

現代日本語の自発音声を収録した『日本語話し言葉コーパス』（CSJ）を用いる。CSJ には講演・自由対話など様々な種類の音声が含まれるが、スタイルによる差を排除するため、「模擬講演」（CSJ のデータ採取のために行われた簡単な講演）のみを使用した。また、方言による差や性別による差を排除するため、「東京で生まれ」「首都圏以外に在住歴がない」「男性」に話者を限定した。

以上の条件を満たす話者は、63 人となった。そのうち、生年代が最も離れた 2 つの集団として、1930 年〜1944 年に生まれた高年層 9 人と、1975 年〜1984 年に生まれた若年層 10 人を取り出し、この 2 つの集団でフォルマント周波数に違いが見られるかを分析した。なお、それぞれの話者は、1 回あたり 10 数分の講演を、平均して約 3 回ずつ行っている。

これらの講演の中から、実際に分析する母音を次のようにして選定する。

・先行子音の影響を排除するため、ポーズの直後にある母音に限定する。CSJ の書き起こしテキストは、0.2 秒のポーズごとに分割されているので、その冒頭のみを対象とする。
・後続子音による影響を小さくするため、「サ・ザ・タ・ダ・ナ・ラ行のいずれかの子音

が続く」母音に限定する。これは、母音や半母音が後続する母音を除外し、さらに、後続する子音の調音位置を統制したものである。

このようにして選定した各母音を対象に、母音区間の中央 1/3 について、F1・F2 の平均値を計測した。計測には Praat（version 6.0.09）を使用し、Formant settings（設定）は Maximum formant (Hz) の値を 5000 として、Number of formants の値は 5 とした[1]。母音が無声化している場合は、声帯の安定した振動が観察されないので、使用不可能と判断して除外した。同様に、後続子音の閉鎖や摩擦が極端に弱まっている場合も、母音区間を特定できないので、使用不可能と判断して除外した。

最終的に、818 個の母音を使用サンプルと位置付けて、分析を行う。高年層と若年層それぞれの使用サンプルの個数は。表 1 の通りである。

表 1　使用サンプルの個数

	ア	イ	ウ	エ	オ	計
高年層’	156	146	24	18	50	394
若年層	184	119	46	18	57	424
計	340	265	70	36	107	818

4. 分析

4.1. サンプルの偏りについて

実際の分析に入る前に、得られたサンプルの偏りについて、3 つの観点から検討する。コーパスからある条件を満たすものを拾い出す、という手法を採っている以上、話者ごとのサンプル数、後続音の音環境、アクセント核の有無、などの条件は、完全には統制できない。そこで、これらの偏りが、世代間比較に際してどの程度影響するかを考察する。

①話者ごとのサンプル数の偏り

例えばアを見ると、話者ごとに見た高年層の平均は 17.3 個であるが、ある話者は 43 個も含む。サンプル数の多い話者の影響で、その世代のフォルマント周波数の分布が、大きく歪められてしまうかもしれない。対応策として、平均の 2 倍以上のサンプルを含む話者については、フォルマント周波数が特異な値を取っていないかを確認する。

②後続音の音環境の偏り

計測する母音のフォルマント周波数は、後続母音によって影響を受けると予想される。これを考慮せずに分析をしてしまうと、後続母音の違いによって生まれるフォルマント周波数の差を、世代間の差異と見なしてしまうかもしれない。後続母音の違い

[1] ただし、オの計測においては、F1 と F2 が接近するため、それらが 1 つのフォルマントとして捉えられてしまうことがあった。これを避けるため、3000Hz までに 3 つのフォルマントが安定的に示されない場合は、Number of formants を 7 に変更した。

を考慮に入れても世代間の差があるのかを検討するために、後続母音については、分散分析の要因として扱う。

③アクセント核の有無による偏り

アクセント核の有無によって、フォルマント周波数に差が出る可能性もある。そこで計測に当たっては各母音にアクセント核が有るかどうかをも入力し、高年層と若年層の世代ごとに、各母音について、アクセント核が有るものと無いものの個数を比較した。両世代とも、アとイに関しては、アクセント核が有るものも無いものも多くのサンプルを含む一方、ウとエとオに関しては、アクセント核のあるサンプルの個数はいずれも一桁であり、無視できるほど少ないことがわかった。したがって、分散分析において、アとイではアクセント核の有無を要因として扱うべきであり、ウとエとオでは要因として扱う必要は無いと判断した。

結果として、①～③の偏りは、世代間比較を不可能にするものでは無いと判断できる。

4.2. 計測値での分析

ここから具体的な分析に入る。まず、計測して得られたフォルマントの数値そのものによる分析を行う。世代ごとに一括した各母音の平均値は図 1 の通りである（描画には NORM: The vowel normalization and plotting suite を使用）。

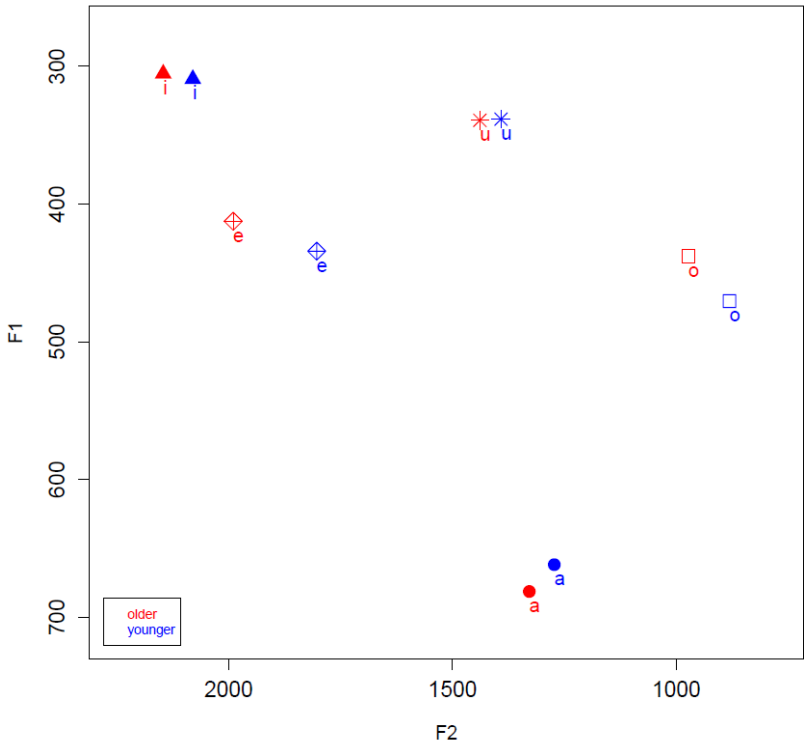


図 1　世代ごとに一括した各母音の平均値

また、世代ごとに区分した全サンプルは図 2 の通りである。

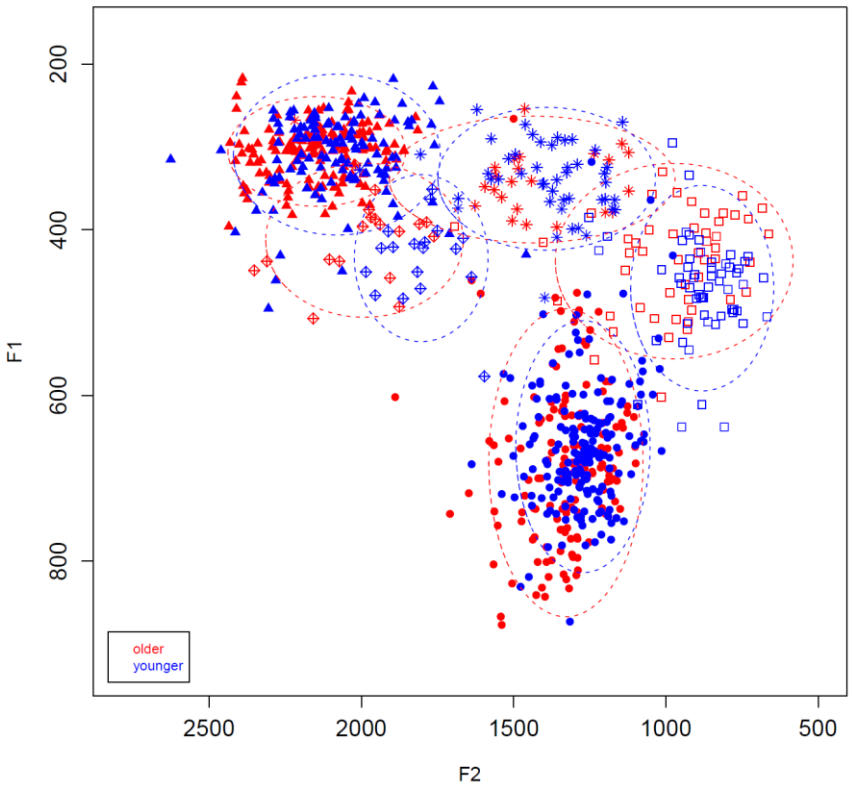


図 2　世代ごとに区分した全サンプル

次に、各母音について分析を行う。アを例として説明すると、まず、各話者のアの平均値について、サンプル数が平均の 2 倍以上である話者が、F1・F2 において最大値または最小値を取っていないかどうかを確かめる（サンプル数の多い話者が極端な値を取っていないことを確認するため）。その上で、「話者ごとのサンプル数の偏りによる影響は少ない」と判断し、各世代ごとに一括して扱って良いものと見なす（結果的には、すべての母音において、このように扱うことが出来ると判断した）。そして、「世代」・「後続母音」・「アクセント核の有無」の 3 つを要因とする分散分析を行う（ウ・エ・オについては、「アクセント核の有無」を考慮する必要がないので、「世代」・「後続母音」の 2 つを要因とする分散分析を行う）。

分散分析は、それぞれの要因が結果に影響を与えているかどうかを調べるものである。本研究においては、「世代」という要因（高年層か若年層か）が、フォルマント周波数の値に影響しているかどうかを調べることが最も重要である。計測値による分析の結果、「世代」という要因による有意差の有無は、表 2 のようになった（分析には ANOVA4 on the web を使用）。有意差があるものについてその大小を説明すると、「アの F1・F2 は、若年層の方が高年層と比べて相対的に小さい」、「イ・エ・オの F2 は、若年層の方が高年層と比べて相対的に小さい」ということになる。

表 2　計測値による分散分析の結果

	ア	イ	ウ	エ	オ
F1	*	n.s.	n.s.	n.s.	n.s.
F2	**	*(**)	n.s.	**	**

※*は危険率 5%で、**は 1%で有意差が示されたもの。n.s.は有意差が無いもの。

※イの F2 は、「3 要因での分散分析（2 要因での分散分析）」の順で結果を併記した[2]。

4.3. 正規化を施した上での分析

計測値での分析に続いて、正規化を施した上での分析を行う。母音の比較における「正規化」とは、男性と女性、あるいは成人と子供など、声道の長さに違いがあると考えられる集団を比較可能にするための、数学的な処理である。フォルマント周波数は声道の長さに影響され、声道が長くなればなるほどフォルマント周波数が小さくなるという関係がある。本研究では話者を男性に限定しているが、4.2.節で示された有意差が、身長差によるものである可能性も、排除できない。そこで、本節では正規化を施した上で同様の分析を行い、それでも有意差が示されるのかを確かめる。

正規化には複数の手法があり、どの手法も長所と短所がある。それぞれの特徴について、Thomas（2011）を参考にした上で、本研究では、「バーク尺度に変換してその差を取る」という手法を採用した。

この手法により得られた数値を用いて、計測値での分析と同様に分析を行う（正規化の計算には NORM を使用）。正規化後の数値による分析の結果、世代による有意差の有無は表 3 のようになった。

表 3　正規化後の数値による分散分析の結果

	ア	イ	ウ	エ	オ
F1’	n.s.	n.s.	—	*	**
F2’	n.s.	*(**)	—	**	n.s.

※イの F2 については、表 2 と同様。

※ウについては、この手法では適切な判断が出来ないため、空欄とした[3]。

[2] 世代（2 種類）、後続母音（5 種類）、アクセント核の有無（2 種類）によってサンプルを分けるので、20 種類（2×5×5）のパターンが必要である。しかし実際には、そのうちの 2 パターン（「高年層・アクセント核有・後続母音 a」と「若年層・アクセント核有・後続母音 i」）に該当するサンプルが 1 個も得られなかった。データがゼロのパターンが存在する場合、分散分析が難しい。そこで次善の策として、後続母音が a と i のサンプルをすべて除外した上での分散分析（3 要因）と、アクセント核の有無という区別を撤廃して上での分散分析（2 要因）を行った。

[3] この手法では F3 の値を使用して正規化が行われる。この手法で正規化を施して 2 つの集団を比較する場合、F3 の値が集団間で大きく異ならないということが前提になっている。本研究の計測時には予備的に F3 も記録したが、ウだけは世代によって F3 の値が大きく異なっていた。したがってこの手法では、ウについては適切に正規化できないと判断した。

5. 結論

計測値による分析（表 2）と正規化後の数値による分析（表 3）から、世代差が示唆されるフォルマントは、①計測値と正規化後の数値との両方で有意差が示されたもの②計測値と正規化後の数値のどちらか一方で有意差が示されたもの、に二分できる。

①イの F2 とエの F2 は、計測値においても正規化後の数値においても、若年層の F2 が小さくなるという傾向が見出された。

②アの F1・F2、エの F1、オの F1・F2 は、計測値あるいは正規化後の数値のどちらか一方で世代差が示されており、世代差が存在する可能性がある。このグループは解釈が難しく、今後の課題としたい。特にオについては、F1 と F2 が接近するため計測が難しく、計測方法自体にも改良の余地があるかもしれない。

以上をまとめると、少なくともイとエの F2 には世代差が存在する可能性が高い。どちらも若年層の方が相対的に小さい値を取っており、イとエが中舌化しつつあると解釈できる。

今後は、話者数を増やしたり、別の要因を設定したり、今回とは異なる正規化の手法を採用したりしても、このような世代差が示されるかを調べる必要がある。また、様々な条件を統制して分析するためには、インタビュー調査を新たに行ったり、これまでに蓄積された録音資料を用いたりすることも有効であろう。

参考文献

【使用データ及び分析に使用したサイト】

国立国語研究所・情報通信研究機構・東京工業大学「日本語話し言葉コーパス」第 5 刷

桐木建始（2002）“ANOVA4 on the Web” https://www.hju.ac.jp/~kiriki/anova4/（2017 年 7 月 31 日最終アクセス）

Thomas, Erik R. and Tyler Kendall（2007）“NORM: The vowel normalization and plotting suite” http://lingtools.uoregon.edu/norm/index.php (accessed July 31, 2017)

【欧文文献・邦文文献】

Thomas, Erik（2011）*Sociophonetics: An introduction.* Basingstoke: Palgrave Macmillan.

井上史雄（1989）「子音の発音の変化」 杉藤美代子（編）『講座日本語と日本語教育 2―日本語の音声・音韻（上）』109-134. 東京：明治書院.

柴田実（2005）「平成の発音変化」『月刊言語』 34(1), 64-72.

杉藤美代子（1997）「日本語音声の音声学的特徴」『BME（生体医工学）』 11(4), 2-6.

P09

母音や摩擦音が後続する撥音の異音について

松井 理直 (大阪保健医療大学)
michinao.matsui@ohsu.ac.jp

キーワード：エレクトロパラトグラフィ (EPG)、撥音、変異音、同化、音節

1. 研究の目的

日本語の撥音は、その音声環境に従って様々な変異音を持つ。特に、破裂口音や破裂鼻音あるいは弾き音といった閉鎖的性質を持つ子音が撥音に後続する場合、その撥音は後続子音の調音位置に関する逆行同化を受けるという点で、全ての先行研究の見解が一致している。しかし、母音や摩擦音といった閉鎖的性質を持たない分節音が後続した撥音がどのような変異を起こすかという点については、必ずしも意見の一致が見られない。

そこで本研究は、母音や摩擦音が後続した撥音の変異について、エレクトロパラトグラフィ (EPG) を用いた調査結果の報告を行う。最初にいくつかの先行研究を概観した後、EPG を用いた実験結果を示す。結論として、撥音に母音が後続した場合には逆行同化が起こるとは限らず、前後の音声環境のうち、より口腔空間の狭い分節音の影響を受ける易いを述べる。また、口腔内の舌運動と口唇の運動が独立した影響を及ぼすことも示す。最後に、この実験結果から分かる日本語の音節構造について、若干の議論を行う。

2. 撥音変異に関する先行研究の概観

2.1 撥音が中線的接触を持つ子音に先行する場合

撥音に破裂口音や破裂鼻音あるいは弾き音といった閉鎖的性質を持つ子音が後続する場合に、撥音が起こす変異については、全ての先行研究で見解が一致している。すなわち、その撥音は後続子音から調音位置の情報を受け取り、逆行同化を起こす。

(1) a. 両唇破裂 [p], [b], [m] が後続する場合：撥音は [m] に変異。
b. 歯茎破裂 [t], [d], [n], [t͡s], [d͡z]，歯茎弾き音 [ɾ] が後続する場合：撥音は [n] に変異。
c. 歯茎硬口蓋破擦音 [t͡ɕ], [d͡ʑ] が後続する場合：撥音は [nʲ] あるいは [ɲ] に変異。
d. 軟口蓋破裂 [k], [g], [ŋ] が後続する場合：撥音は [ŋ] に変異。

ただし、ラ行子音が後続する場合については、この子音自体が様々な変異音を持つことを反映して、意見が一致するとは限らない。例えば、Vance (2008) は [r] 音に先行する撥音は鼻音化接近音 [ɹ̃], [ĩ] に変異するという。また 益子 (2009) は、ラ行子音の異音として [l] 音も考慮に入れ、この音が撥音に後続する場合には [l̃] になると指摘している。接近音や側面接近音が調音方法の一種であることを考慮すると、これらの指摘は非常に興味深い。撥音が調音位置の逆行同化のみならず、ある種の調音方法に関する情報の逆行同化も受けることを示唆しているからである。こうしたラ行子音の問題を置いておくと、破裂音・破裂鼻音・弾き音そして歯茎側面音はいずれも中線的接触を持つ調音であり、これらの子音が撥音に後続する場合は、少なくとも調音位置に関する拘束変異が起こるとまとめておいてよいだろう。

2.2 撥音が末尾に位置する場合

これに対し、撥音が末尾に位置しており、他の分節音が後続しない環境では、撥音の変異について見解が食い違う。例えば、藤村・大泉 (1972) は [ɴ] に変異するとし、斎藤 (2006) は「本」や「缶」では [ɴ] だが、「銀」や「千」では [ŋ] としている。また、村木・中岡 (1989) は「口蓋垂と奥舌面の間が狭まるが、口腔内の閉鎖がない継続鼻音」と述べており、これは [ũ̠] か [ũq̠] に相当する音と見てよいだろう。同様に Vance (2008) も、撥音変異に接近音の鼻音化 [ũq̠] があり得ることを指摘している。

末尾の撥音に関しては、吐師ほか (2016) による生理学的手法を用いた精密な研究が極めて興味深い。彼らは、X 線マイクロビームを用いた実験により、末尾に位置する撥音変異が [ɴ] に限定されず、自由度が高く、話者間変動も大きいことを見いだした。先行研究における記述のばらつきも考慮すると、末尾に位置する撥音は自由変異であり、その自由度は、Itô & Mester (1995) が指摘した「撥音の調音位置は過小指定されている」という性質と関係している可能性がある。ただし、超音波測定を用いて撥音変異を実験的に調べた Yamane (2013) によると、こうした撥音変異の自由変異も、軟口蓋に明確な調音位置の指定を持っている [k] 音に比べて自由度が高いわけではないという。この指摘は、撥音の性質を考える上で見逃せないものの 1 つである。

2.3 撥音が摩擦子音に先行する場合

撥音に摩擦音が後続する環境における変異については、益子 (2009) の記述が非常に詳しい。益子は後続摩擦音に従って鼻母音に変異すると見ており、[s], [ɕ], [z], [ʑ] が後続する場合は [ĩ̠] に、[ç] が後続する環境では [ĩ] に、[ɸ] が後続する時には [ũ̠] に変異するとしている。興味深いのは [h] 音に先行する撥音の変異で、これは [h] 音に後続する母音に従って変異音が決まるという。声門摩擦音は口腔の調音が指定されておらず、[h] 音の調音時点で口腔の構えは後続母音と同一のものになるため、これも通常の逆行同化とほぼ同様の現象であると見てよい。

摩擦音に先行する環境で撥音が鼻母音に変異するという点は、多くの先行研究が一致するところであるが、微妙な点で相違も見られる。服部 (1951) は、摩擦音に先行する撥音は後続摩擦音に同化し、鼻音化摩擦音になる (ただしその摩擦的噪音がなくなり、鼻母音になることもある) と指摘した。同じく Kochetov (2014) も、[s] 音に先行する撥音変異を EPG によって測定し、[ĩ] か [z̃] に変異すると述べている。これらの現象は、前述した [Ĩ] 音への同化とと同様、撥音は単に調音位置の逆行同化を受けるだけでなく、調音方法の逆行同化も受けることを示しているのだろう。

しかし村木・中岡 (1989) によれば、摩擦音に先行する撥音変異はもう少し複雑なようだ。彼らは、撥音に摩擦音が後続した場合、「口構えは前後の音に従い、舌は口蓋に向かって持ち上げられるが、口腔内の閉鎖がない継続鼻音 (すなわち鼻音化摩擦音か鼻母音)」に変異するという。「前後の音に従い」という点に注目されたい。すなわち彼らによれば、撥音は逆行同化のみならず、順行同化の影響を受ける可能性も持つ。これは、次節で述べる母音が後続する環境における撥音変異でも同様に問題となる。

2.4 撥音が母音に先行する場合

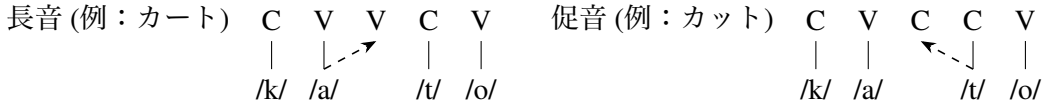
母音が撥音に後続する場合、ほぼ全ての研究が撥音は鼻母音に変異するとしており、この点では違いがない。大きく異なるのは、順行同化か逆行同化かという点である。服部 (1951) や 益子 (2009) は、撥音が後続母音に従って変異すると見なす。斎藤 (2006) は撥音変異の全体的な説明として「後続音に同化」と明記しているが、具体例を見ると、「線を」が [seẽ.o]

となっており、例は少ないものの順行同化もあり得ると考えているようだ。これに対し、村木・中岡 (1989) は「口構えは前後の音に従い、舌は口蓋に向かって持ち上げられるが、口腔内の閉鎖かがない鼻音 (継続鼻音すなわち鼻音化摩擦音か鼻母音)」と述べ、「全員 [zeẽiĩ]」「禁煙 [kiĩeẽ]」、「安易 [aãi]」といった発話のスペクトログラムを示して、順行同化の例を多く挙げている。

この点で注目すべきは、Vance (2008) であろう。彼は、撥音が母音・接近音・流音・摩擦音の前で鼻母音に変異すると共に、環境や話者によって先行母音に同化したり、「調音上さらに狭窄の狭い音」に自由変異する可能性について指摘した。次節で述べる実験結果から、本研究も「調音動態としてより狭窄が狭い」という条件こそが撥音変異の本質であると考える。

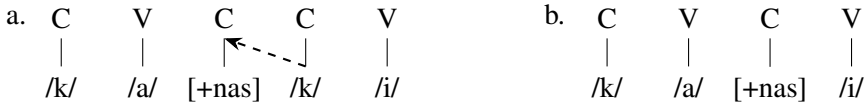
2.5 音節構造内のスロット表示を巡って

撥音が逆行同化の影響のみを受けるのか、あるいは順行同化も影響するのかという問題は、音韻論における撥音のスロット表示とも深い関係を持つ。一般に、日本語の長音や促音は Obligatory Contour Principle (OCP, McCarthy 1986) に違反しない形で、



のような構造を持つ。では、同じ特殊拍である撥音はどのような表示を持つのだろう。「ミント」に見られる逆行同化は、撥音が基底から [+nas] 素性を持っているという点を除き、基本的に促音と並行的な関係なので、(2) のような構造を持つと考えてよいかもしれない。すなわち、「官記」で逆行同化が起こるのは撥音 ([+nas] 素性) が C スロットに結びついており、かつ [cont] 素性や調音位置に関する素性を持っていないため、後続の C スロットに結びついている [cont] 素性や調音位置素性が拡張される。しかし、「官位」のように撥音の前後が母音だった場合、撥音の “C” するっとは同化を引き起こす力にならない。そうなると、何が同化を引き起こすのだろうか。本研究の目的は、こうした問題も明らかにするために、母音や摩擦音が後続する場合の撥音変異の詳細を EPG を用いて調査する点にある。

(2) 撥音 (例：官記) の構造　　「官位」の構造？



3. 摩擦音／母音に先行する撥音変異に関する実験

3.1 実験方法と分析方法

実験は被験者 4 名 (東京方言女性 1 名、関西方言女性 1 名、関西方言男性 2 名) に対し、EPG および鼻腔用マイクと口腔用マイクを装着させ、3 モーラの刺激語をランダムに 5 回ずつ発話させる形で行った。本論文で分析の対象とする刺激語のパターンは以下の通りである (これ以外のデータも採ったが、紙面の関係上、本稿では述べない)。有意味語と無意味語が混ざっているため、EPG を装着する前に事前に全ての刺激語を順に 5 回連続で音読させ、無意味語に慣れてもらった。またアクセントは平板型で読むように指示しているが、起伏型で読まれた場合に言い直し等はさせていない。また、正中面からビデオカメラによって口唇の動きを撮影した。

(3) a. 母音–撥音–母音：「あんい」「あんう」...「おんお」

b. /k/–母音–撥音–母音：「かんい」「かんう」...「こんお」

c. 母音–撥音–/h/–母音：「あんは」「あんし」...「おんほ」

d. /k/–母音–撥音–/h/–母音：「かんは」「かんし」...「かんほ」

撥音が母音に挟まれている環境では、いずれの母音も鼻音化する可能性があるため、撥音区間の認定は、鼻腔用マイクおよび口腔用マイクから収録した音とスペクトログラムを手がかりに行った。なお、撥音が母音に挟まれている場合でも、図 refN-inter (a) の「かんた」と類似した撥音区間を持つ図 refN-inter (b) のようなパターンは、撥音部が鼻子音 [ŋ], [ɴ] であると見なし、図 refN-inter (C) のように撥音部であっても口腔用マイクが完全に音声を拾っている場合を鼻母音であると見なす。

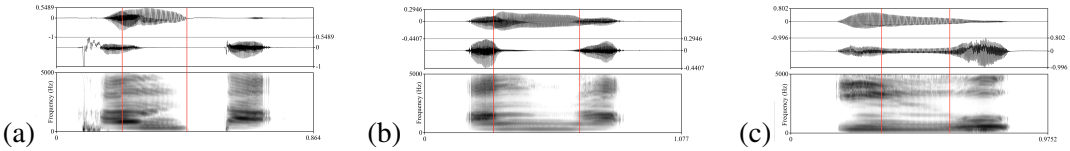


図 1　撥音の認定 (a) かんた　(b) あんあ　(c) いんう

EPG データは、4 名の被験者に対し人工口蓋床を各被験者ごとに用意し、サンプリング周期 10 ms で収集した。この人工口蓋床は山本一郎氏によって調音への負担が最小限になるように開発されたもので、異なる話者の調音を比較できるよう標準化され、前後方向に歯茎 2 列、後部歯茎 2 列、硬口蓋 3 列、軟口蓋境界部 1 列の電極配置を持つ。また、左右方向は歯茎最前列のみ電極が 6 点、他の列は電極が 8 点配置され、歯茎から軟口蓋境界部まで計 62 点の計測が可能となっている。なお、EPG データの収録には WinEPG (Articulate Instruments Ltd.) およびタブレット端末を用いて行った。

3.2　分析結果

分析例として、「あんい・いんあ」「あんう・うんあ」のスペクトログラムと EPG 遷移の結果を図 2 に示す。こうした広母音・狭母音の組

う違い

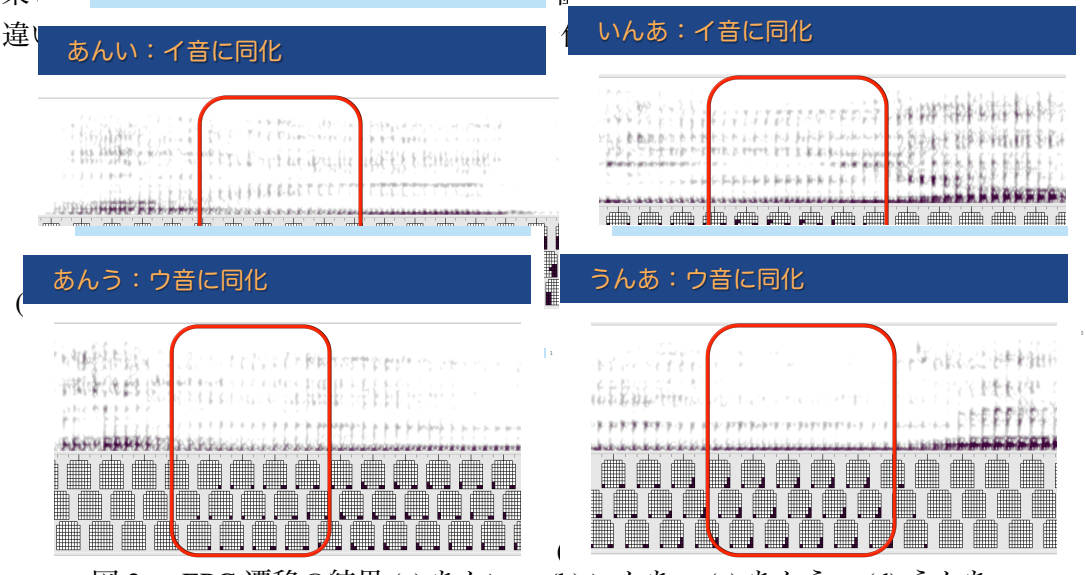


図 2　EPG 遷移の結果 (a) あんい　(b) いんあ　(c) あんう　(d) うんあ

撥音に先行／後続する母音環境と撥音部の鼻母音変異の結果を、表 1 にまとめておく。基本的に、撥音の前後にある母音のうち、口腔空間のより狭い母音が鼻母音化していることが分かる。また、/i/, /u/ という狭母音に撥音が挟まれた場合は、基本的に [ĩ] に変異する傾向が強い。これは、/i/ 音が palatal という性質を明確に持つのに対し、/u/ 音は強い素性を持たないことに起因しているように思われる。これに対し、/e/, /o/ という半狭母音に撥音が挟まれた場合は、先行母音に引きずられた鼻音に変異しやすい。この原因は明確ではないが、音節境界を越える逆行同化よりも、音節内で起こる順行同化を好む可能性が考えられる。この点については、今後の課題としたい。なお表中の “?” は、EPG の反応が取れないために明確なことがいえないことを示す。また、「母音–撥音–ハ行子音–母音」の連鎖についても、表 1 とほぼ同一の結果が得られた。これはハ行子音の調音動態から言っても当然の結果であろう。

表 1　撥音部における鼻母音の音価

		後続母音				
		[ɑ]	[i]	[ɯ]	[e]	[o]
先行母音	[ɑ]	?	[ĩ]	[ɯ̃]	[ẽ]	?
	[i]	[ĩ]	[ĩ]	[ĩ]	[ĩ]	[ĩ]
	[ɯ]	[ɯ̃]	[ĩ]	[ɯ̃]	[ɯ̃]	[ɯ̃]
	[e]	[ẽ]	[ĩ]	[ɯ̃]	[ẽ]	[ẽ]
	[o]	NC	[ĩ]	[ɯ̃]	[õ]	?

ハ行子音に対し、「母音–撥音–/s/–母音」「/k/–母音–撥音–/s/–母音」の連鎖では Kochetov (2014) の結果とほぼ同一の結果が得られ、服部 (1951) が指摘するように撥音部は [s̃], ɕ̃ に変異していた。その一例を図 3 に示す。このことは、破裂音と同じく /s/ 音が後続する場合にも調音位置の逆行同化が起こっていることを示す。この性質は[redacted]という Itô & Mester (1995) の指摘と

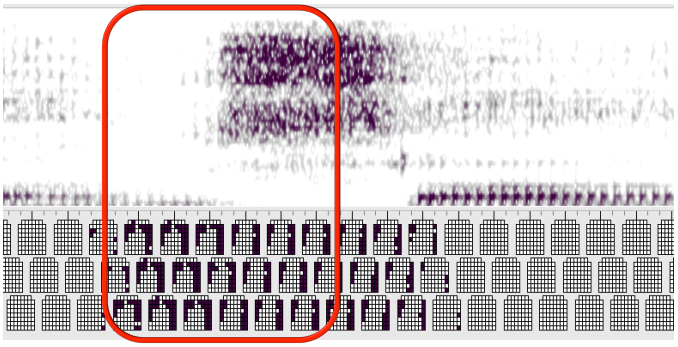


図 3　「あんさ」における EPG 遷移の結果

4.　総合論議

EPG を用いた撥音変異の実験から、撥音の変異は前後に存在する音韻環境のうちで口腔空間のより狭いものに強く影響されるということが明らかとなった。撥音に子音が後続する時に逆行同化を起こすのも、日本語の phonotactics として撥音に先行する音は必ず母音であり、撥音に後続する子音のほうが必然的に口腔空間がより狭いことが原因であると考えられるだろう。すなわち、撥音の逆行同化は 2.5 節で議論したような CC-slot 連鎖が引き起こす

ものではない。このことは、促音の逆行同化の再検討にもつながるだろう。すなわち、促音における逆行同化も CC-slot 連鎖が引き起こすものではない可能性がある。促音も撥音と同じく「口腔空間のより狭い環境に影響される」とすれば、促音の場合は (語末促音を除き) 必ず後続環境が子音であるため、必ず逆行同化が起こることも不思議ではない。

この議論は、日本語の音韻表示において撥音や促音に C-slot が必要かという問題につながる。「口腔空間のより狭い環境に影響される」という性質は、音節末という位置が要求する sonority の低さの反映と考えることも可能であろう。この場合、C-slot は不要となり、撥音が「鼻母音」に変異することとも自然に説明がつく。一方、C-slot を仮定する場合、鼻母音への変異は 'C' を「プロトタイプカテゴリー」と考えない限り、矛盾が起こる。この点については、促音が基底で featureless なのかといった問題とも絡むため、また稿を改めて議論を行いたい。

謝辞：本研究を実施するにあたり、人工口蓋床を作成してくださった山本一郎先生に深く感謝いたします。なお、本研究は科学研究費・基盤研究 (C)「音声知覚における摩擦性極周波数特性の影響に関する総合的研究」の援助を受けました。

参考文献

藤村 靖・大泉 充郎 (1972) 『音声科学』 東京大学出版会. 東京.

Itô, J., & Mester, A. (1995) "Japanese phonology." In J. A. Goldsmith (Ed.), *The handbook of phonological theory*. (pp. 817–838). Blackwell. Oxford.

Kochetov, A. (2014). *Japanese in the typology of nasal place assimilation: Electropalatographic evidence*. (Formal Approaches to Japanese Linguistics (FAJL) 7, Handout).

益子 幸江 (2009) 「音声記号」 今泉 敏（編）『言語聴覚士のための音声学・言語学』20–36, 医学書院. 東京.

McCarthy, J. J. (1986) "OCP effects: Gemination and antigemination." *Linguistic Inquiry*, 17, 207–263.

村木 正武・中岡 典子 (1989) 「撥音と促音—英語・中国語話者の発音—」 杉藤 美代子（編）『講座 日本語と日本語教育第 3 巻』139–177, 明治書院. 東京.

斎藤 純男 (2006) 『日本語音声学入門』 三省堂. 東京.

服部 四郎 (1951) 『音声学』 岩波書店. 東京.

吐師 道子・小玉 明菜・三浦 貴生・大門 正太郎・高倉祐樹・林 良子 (2016) 「日本語語尾撥音の調音実態：X 線マイクロビーム日本語発話データベースを用いて」 『音声研究』 18: 2, 95–105.

Vance, T. J. (2008) *The sound of Japanese*. Cambridge University Press. New York.

Yamane, N. (2013). *'Placelessss' Consonants in Japanese: An Ultrasound Investigation*. Unpublished doctoral dissertation, The Uuniversity of British Columbia, type = Ph. D. dissertation,.

P10

日本語を母語とする英語学習者におけるシュワーの削除に関する研究

坂本 洋子（獨協医科大学）

y-saka@dokkyomed.ac.jp

1. はじめに

英語母語話者の発話において、シュワー（ə）の削除が起こることが報告されている。先行研究によると、シュワーの削除を引き起こす条件として、強勢の位置（強勢の前か、後か)、語彙頻度、形態的複雑性、発話速度、スピーチスタイル、話者の性別、方言などが提案されている（Dalby, 1986; Patterson et al, 2003)。これらの条件の内、発話速度に関しては、通常の速度よりも速い速度の方がシュワーの削除の頻度が高くなり、速い速度では 35%から 45%の削除率であったと報告されている。また、強勢の位置は強弱弱の 3 音節から成る語がシュワーの削除率が最も高いことが示されている。それでは日本語母語話者は英語を発話する際に、シュワーの削除に関して、英語母語話者と同様の傾向を示すのであろうか。本研究では、日本語を母語とする英語学習者を対象として、単語の読み上げと文の読み上げを行う 2 つの実験を行った。

2. 実験 1

実験 1 では日本語を母語とする英語学習者における単語内のシュワーの削除を調べることを目的とする。分析対象はシュワーの削除が生じる可能性のある語を 2 種類の速度で発話した音声である。

2.1. 実験方法

2.1.1. 言語材料

本研究ではシュワーの削除が起こる可能性のある単語 53 語をテスト語とした。シュワーの削除が起こる可能性のある単語は Patterson et al. (2003)で用いられた単語のうち、日本語を母語とする大学生に馴染み度の高い単語を用いた。単語の種類は音節数と強勢の位置に基づく 3 種類である。強弱のリズムを持つ 2 音節語（2PRE)、弱強弱のリズムを持つ 3 音節語（3PRE)、強弱弱のリズムを持つ 3 音節語（3POST）である。単語の例はそれぞれ、select /səlékt/ (2PRE), solution /səlú:ʃən/(3PRE), avenue /ǽvən(j)ù:/ (3POST)である。

2.1.2. 被験者

被験者は日本語を母語とする大学生 16 名であった。男性は 6 名、女性は 10 名であった。学年は 1 年生から 4 年生までであり、専攻は英語 13 名、交流文化学科 2 名、フランス語 1 名であった。学年は 1 年生から 4 年生までであり、平均年齢は 20 才であった。出身は日本であり、英語圏の海外滞在経験がない被験者は 6 名、英語圏の海外滞在経験がある被験者

は 10 名であり、期間は 2 週間から 1 年であった。TOEIC の点数は 400 点から 810 点であり、平均は 738 点であった。

2.1.3. 実験手順

実験手順は Dalby (1986)と氏平・窪薗(2000)の方法を参考に、以下の手順で行った。被験者は 1 人ずつ静かな部屋で実験に参加した。被験者はまずテスト語 53 語が書かれた紙を渡され、単語に慣れるように指示された。その際制限時間は設けなかった。ランダムに配置した 53 語のテスト語を一まとめにして 1 回とし、テスト語の読み上げを各被験者について 5 回行い、IC レコーダーを使って録音した。そしてこの 5 回の録音のうち第 2 回目から第 5 回目までの 4 回のテスト語の録音を測定の対象とした。読み上げの速度は普通の速度と意味が取れる範囲でできるだけ速い速度とを 1 回ごとに交互に行うように指示された。

シュワーの削除が起こっているかどうかの判断は、Praat を用いて行った。シュワーの削除は音声的な現象であり、段階的(gradient)であるが、今回は削除が完全に起こっている場合のみを扱うことにした。シュワーの削除の判断基準は、聴覚による確認と波形の視覚による確認を行い、シュワーが完全に消えているものとした。これは Dalby (1986)と同じ基準である。

2.2. 実験結果

16 名の被験者の音声データを分析し、シュワーが削除されているテスト語の数を数えた。回数ごとのテスト語数は 53 語×16 名で合計 848 語であった。アクセントや発音を間違えたものは誤答とした。16 名の被験者の録音回数ごとのシュワーの削除数を表 1 に示す。

表 1：テスト語読み上げにおけるシュワーの削除数と誤答数

	シュワーの削除数	誤答数
練習	5	22
普通 1	6	17
速い 2	6	17
普通 3	4	19
速い 4	9	16

2.3. 議論

実験 1 の結果より、テスト語の読み上げにおいて、日本語母語話者はシュワーの削除率が最も高い速い 4 で 1%、普通 3 で 0.4%であり、ほとんどシュワーの削除を行っていなかった。また発話速度による違いや、音節数と強勢の位置による単語の種類による違いは見られなかった。このことは英語母語話者と日本語母語話者では異なるメカニズムでシュワーの削除を行っている可能性を示唆するものであると考えられる。

3. 実験 2

実験 2 では、日本語母語話者における文中のシュワーの削除を調べることを目的とする。分析対象は、シュワーの削除が生じる可能性のある語を組み込んだキャリア文を 2 種類の速度で発話した音声である。

3.1. 実験方法

3.1.1. 言語材料

実験 1 でテスト語として用いられた単語 53 語を埋め込んだキャリア文 53 文を言語材料とする。シュワーの削除が起こる可能性のある単語は Patterson et al. (2003)で用いられた単語のうち、日本語を母語とする大学生に馴染み度の高い単語を用いた。単語の種類は音節数と強勢の位置に基づく 3 種類である。強弱のリズムを持つ 2 音節語（2PRE）、弱強弱のリズムを持つ 3 音節語（3PRE）、強弱弱のリズムを持つ 3 音節語（3POST）である。単語の例はそれぞれ、select /səlékt/ (2PRE), solution /səlú:ʃən/(3PRE), avenue /ǽvən(j)ù:/ (3POST)である。キャリア文は『SCN 小学館コーパスネットワーク』の British National Corpus から検索した文を、英語母語話者が確認した上で使用した。キャリア文は日本語を母語とする大学生に馴染み度の高い語で構成されている。

3.1.2. 被験者

被験者は実験 1 に参加した日本語を母語とする大学生 16 名である。実験 1 が終わった後すぐに実験 2 に参加した。

3.1.3. 実験手順

実験手順は基本的には実験 1 と同じである。被験者はまずキャリア文 53 文が書かれた紙を渡され、文に慣れるように指示された。その際制限時間は設けなかった。ランダムに配置した 53 文のキャリア文を一まとめにして 1 回とし、キャリア文の読み上げを各被験者について 5 回行い、IC レコーダーを使って録音した。そしてこの 5 回の録音のうち第 2 回目から第 5 回目までの 4 回のキャリア文の録音を測定の対象とした。読み上げの速度は普通の速度と意味が取れる範囲でできるだけ速い速度とを 1 回ごとに交互に行うように指示された。

シュワーの削除が起こっているかどうかの判断は、Praat を用いて行った。シュワーの削除は音声的な現象であり、段階的(gradient)であるが、今回は削除が完全に起こっている場合のみを扱うことにした。シュワーの削除の判断基準は、聴覚による確認と波形の視覚による確認を行い、シュワーが完全に消えているものとした。これは Dalby (1986)と同じ基準である。

3.2. 実験結果

16 名の被験者の音声データを分析し、シュワーが削除されているテスト語の数を数えた。回数ごとのテスト語数は 53 語×16 名で合計 848 語であった。アクセントや発音を間違えた

ものは誤答とした。16 名の被験者の録音回数ごとのシュワーの削除数を表 2 に示す。

表 2：キャリア文読み上げにおけるシュワーの削除数と誤答数

	シュワーの削除数	誤答数
練習	11	18
普通 1	15	20
速い 2	13	22
普通 3	13	16
速い 4	13	20

3.3. 議論

実験 1 と比較するとキャリア文読み上げにおいて、やや削除数の増加が見られた。しかしながら、最も削除数が多い普通 1 の 15 でも削除率としては 1%であり、Dalby (1986)による英語母語話者の文の読み上げにおける普通の速度の削除率 6%と比べて少ない傾向が見られた。さらに先行研究では、英語母語話者は速い速度での文の読み上げでは削除率がさらに 43%と高くなる傾向が見られたが、今回の結果は異なる傾向を示している。このことは、英語母語話者は、文の読み上げにおいて、速い速度ではよりリラックスした、カジュアルな話し方になるのに対し、日本語母語話者は速い速度であっても、模範的な発音を行おうと注意深く発音していた可能性があると考えられる。また、日本語母語話者は速い速度の時であっても、英語母語話者の場合ほど、速度の差が出ていなかった可能性が示唆される。

4. 結論

本研究の結果から、日本語を母語とする英語学習者は、英単語と英文の読み上げ時にシュワーの削除をほとんど行っておらず、発話速度や強勢の位置による差は見られないことが明らかになった。ゆえに、シュワーの削除に関して、日本語母語話者は英語母語話者とは異なるメカニズムを持つ可能性が示唆されると考えられる。

参考文献

Dalby, J. (1986). “*Phonetic structure of fast speech in American English*,” Indiana University Linsuist Club, Bloomington.

Patterson, D., LoCasto, P. C and Connine, C. M. (2003). “Corpora analyses of frequency of schwa deletion in conversational American English,” *Phonetica* 60, 45-69.

氏平明・窪薗晴夫(2000)「日本語における母音長の中和について」『第 14 回日本音声学会全国大会予稿集』133‐138.

SCN 小学館コーパスネットワーク http://www.corpora.jp/

P11

神奈川県三浦市方言における老年層話者のアクセントについて

坂本 薫（國學院大學大学院生）
lukun84@icloud.com

１．はじめに

本発表では、対面調査によって得られた神奈川県三浦市方言の老年層話者のアクセントについて述べる。神奈川県の方言は、俚言の蒐集と方言区画の設定に関しては豊富な先行研究があるが、音声、音韻の面ではまだその実態を明らかにする余地が残されている。坂本（2017）では、三浦市方言の老年層話者の発音の観察から、伝統的な音声・音韻的特徴について言及した。その結果を受けて、アクセントにおいても、当該方言の古相の保持が期待されたため、同一の話者を対象に調査を続行している。本発表は調査の結果から、アクセントの体系と名詞と動詞、形容詞の活用形のアクセントについて述べる。

２．三浦市の位置

２．１地理的概況と産業、交通

神奈川県東南部の三浦半島の最南端に位置する、人口約 4 万 5 千人の市である。北部を横須賀市と接している。北部は農村地帯で畑作が主で、南部沿岸は江戸時代より漁港として発達。江戸時代は天領。1955（昭和 30）年、二町一村が合併、三浦市として市制施行。今回の調査地点である三浦市三崎は、三浦半島の南西部、相模湾沿岸に位置する。

鉄道が三浦市内まで延伸したのは 1966（昭和 41）年。1975（昭和 50）年、市内北西部に三崎口駅が開業するが、市の中央部には到達しておらず、以南の交通はバス路線が主となる。京浜急行線で品川駅～三崎口駅間の所要時間はおよそ 65 分である。

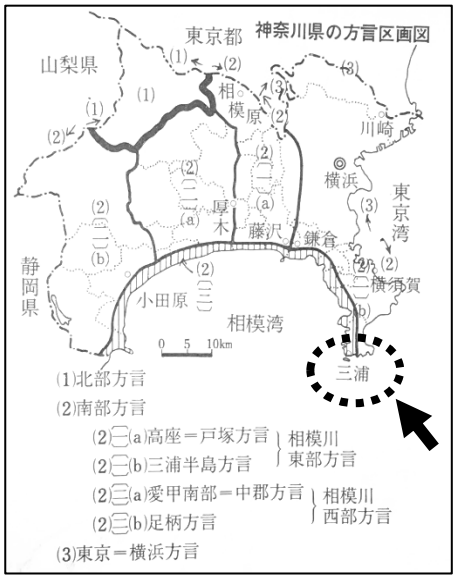


図 1「神奈川県方言」『現代日本語方言大辞典』より。矢印及び点線は筆者による。

２．２方言区画上の位置

全国方言の区画における位置づけとして、東條操（1954）によると、神奈川県方言は東部方言に含まれる関東方言の、さらに西関東方言に含まれると述べられている。

神奈川県内の方言区画は、日野資純（1952、1964）等で述べられている（図 1 ）。三浦市方言は、図中（２）の南部方言に含まれる相模川東部方言の下位区分である三浦半島方言に属する。ただし、西部沿岸部は相模湾沿岸方言の区画にも含まれる。

坂本（2017）では、三浦市で生育した話者を対象にした調査から、三浦半島方言の音声的特徴である[ʒe]と、相模湾沿岸方言の音声的特徴である[je]、明治期東京のアクセントと共通する「～ヤガレ」形式のアクセントが保持されていることについて言及した。

３．調査について

調査は 2017 年 3 月～ 7 月、三浦市内で行った。話者は三浦市南部、旧三浦郡二町谷で生まれ育った 81 歳男性（調査当時）。現役時代は市内で会社員として勤務。外住歴は無い。両親ともに現三浦市域で生まれ育っている。

調査は対面式で、調査項目は名詞、動詞、形容詞とその活用形。名詞語単独と「が」「を」「に」等を付けた、もしくは助詞が脱落した文の発音を観察した。動詞と形容詞については、語単独の発音と、活用形を含む文を作成してもらい、発音してもらった。調査対象は平山輝男（1960）所収の「全国アクセント比較表」に掲載されている語が中心で、これに多拍語を加えた。

４．調査結果と考察

４．１方言的特徴の概観[1]

調査により観察された方言的特徴を挙げる。語中語末のガ行音は規則的に鼻濁音[ŋ]が観察された。また、無声子音に挟まれた狭い母音の無声化は頻繁に生じるが、それに伴うアクセントの下がり目の移動は規則的ではなく、たとえば「吹く」[ɸɯ̥˥kɯ]のように無声化した拍の後に下がり目が来る発音もしばしば観察された。連母音は「ムネータ（胸板）」や「カンゲール（考える）」、「カー（買う）」、「アエー（青い）」など、ai、ae、au、oi には融合が観察されたが、「アツイ（熱い）」「ウスイ（薄い）」など ui の融合は観察されず「カイー（痒い）」などごく少数の語にしか見られなかった。このほか、ヒ→シの交替「シッパル（引っ張る）」や、動詞語末のラ行の撥音化「トンソーダ（取るそうだ）」に加え、「コイ（これ）」「ソイ（それ）」や「デライタ（出られた）」など、一部の語にレ→イの交替が観察される。

４．２アクセントの体系

東京語のアクセント同様、アクセントの「下がり目」の有無、そしてその位置が弁別的特徴であり、その対立は n 拍の語に対して n＋ 1 種ある。表 1 に 3 拍語までのアクセントの体系を示した。

表 1：三浦市方言のアクセントの体系

拍	音韻論的解釈	語例
1	/○/	柄、蚊、毛、子、血…
	/○˥/	絵、尾、木、酢、田、手、荷…
2	/○○/	飴、蟻、烏賊、梅、釘、行く、着る…
	/○○˥/	石、歌、北、寺、足、鍵…
	/○˥○/	斧、秋、汗、婿、夜、会う、見る、良い…
3	/○○○/	欠伸、錨、鎖、衣、魚、机、遊ぶ、借りる…
	/○○○˥/	小豆、力、頭、扇、鏡、刀、瓦…
	/○○˥○/	五つ、さざえ、油、余る、起きる、赤い、黒い…
	/○˥○○/	朝日、嵐、もみじ、帰る、通る…

４．３名詞のアクセント

１～３拍の語のアクセントを、類別語例を中心に表２に示した。表中の 1，2 拍の名詞に注目

[1] 音声の表記は片仮名と一部 IPA を用いた。アクセントの下がり目は記号「˥」で示した。以降のページも同様。

すると、たとえば1拍Ⅰ類の「帆」や、2拍Ⅲ類の「神」、「雲」で（この2語は頭高型のアクセントも観察されたが）、それぞれ伝統的なアクセントが保持されているため、語の類で見たときに例外が少ないように見える。なお、尾高型の名詞に助詞「の」が接続する際、東京のアクセントは「ハナノイロ┐ガ…（花の色が…）」のように平板に発音されると言われるが、今回の調査では言われている通り平板に発音される場合もあったが、「ウラ┐ノハタケデ…（裏の畑で…）」のように下がり目が観察される場合もあった。これは3拍以上の名詞についても同様である。

3拍の名詞については、尾高型のアクセントと中高型のアクセントに特徴が見える。尾高型のアクセントについて、例えばⅡ類「小豆」、「毛抜き」、Ⅳ類「いたち」、「林」、「筵」などは東京のアクセントの古相と共通する。「とかげ」、「くらげ」に関しては方言独特のアクセントの可能性がある。中高型のアクセントが多く観察される点については日野資純（1984）などでも指摘されている通り、神奈川県方言の一つの特徴であると言える。Ⅲ類の「小麦」や「栄螺」、Ⅴ類の「油」や「心」など、東京のアクセントの古相と共通するものもあれば、「茸」「マミヤ（眉）」など、地域的特色のあるアクセントと考えられる語もある。ちなみに、神奈川県方言の特色として、中高型の「南瓜」のアクセントが知られているが、話者は方言としては俚言の「ト┐ーナス」を使用しており、語形「カボチャ」は後から学習した語であったためか共通語と同じ平板型のアクセントであった。

表2：三浦市方言老年層話者　1～3拍名詞のアクセント

〈 〉内の数字は類。表中*は頭高型も。**は尾高型も。＊＊＊は平板型も。

拍	型	語例
1	平板型	柄、蚊、毛、子、血、帆…〈Ⅰ〉名、葉、日、藻〈Ⅱ〉
	起伏型	世〈Ⅰ〉 矢〈Ⅱ〉 絵、尾、木、酢…〈3〉
2	平板型	飴、蟻、烏賊、牛…〈Ⅰ〉
	尾高型	石、岩、歌、音、北…〈Ⅱ〉足、網、泡、家、神*、雲*、坂*…〈Ⅲ〉
	頭高型	斧〈Ⅲ〉 跡、粟、息、糸…〈Ⅳ〉 秋、汗、雨、鮎…〈Ⅴ〉
3	平板型	欠伸、筏、錨、田舎、鰯…〈Ⅰ〉間、桜、扉、百足 〈Ⅱ〉 黄金、二十歳〈Ⅲ〉 畑、東**〈Ⅳ〉 兎、鰻、狐、雀、背中…〈Ⅵ〉 後ろ、鯨、薬、盥〈Ⅶ〉
	尾高型	麓***〈Ⅰ〉 小豆、毛抜き、とかげ、釣瓶、二つ、二人、夕べ、力〈Ⅱ〉 頭、いたち、瓦、拳、鏡、刀、林、筵…〈Ⅳ〉 くらげ、豆腐、ムジナ（狸）
	中高型	小麦、栄螺〈Ⅲ〉 五つ、思い、境〈Ⅳ〉 油、命＊、心、姿*、涙*、柱、枕*、まなこ*〈Ⅴ〉 茸、コケラ（鱗）、ひじき、マミヤ（眉）、若芽、山葵
	頭高型	えくぼ、緑〈Ⅱ〉 岬〈Ⅲ〉嵐、紅葉、わらび〈Ⅳ〉 朝日、鰈***、きゅうり、ざくろ〈Ⅴ〉 烏、高さ〈Ⅵ〉 蚕、兜、便り、病〈Ⅶ〉

次に、4拍以上の語について、目立った特徴を以下に挙げる。「足音」や「雷」など、尾高型から中二高型への変化が指摘されている、いわゆる「A型B型のアクセント」については、調査した語のほとんどで尾高型のアクセントが観察された。このほか、複合名詞について、「サケカ┐ス（酒粕）」や「ガラスマ┐ド（ガラス窓）」など、後部要素の下がり目が生かされるアクセントや、

「アꜛカトンボ（赤とんぼ）」や「ウꜛミボーズ（海坊主）」など、前部要素の下がり目が活かされる頭高型のアクセントなど、東京および周辺域における多拍の名詞の伝統的な姿が観察された。

４．４動詞のアクセント

動詞は、平板型と起伏型のアクセントの対立が明確にある。2 拍の語では「イク（行く）」と「カꜛク（書く）」の 2 種、3 拍の語では「アガル（上がる）」、「アマꜛル（余る）」、「トꜛース（通す）」の 3 種の方の対立がある。本稿ではまず、3 拍の動詞の活用形のアクセントについて主に述べる。表３には「遊ぶ」、「入る」、「選ぶ」、「帰る」の活用形のアクセントが示してある。Ⅲ類の動詞「入る」は、東京のアクセントは頭高型で、活用形のアクセントも頭高型の動詞と同じになるが、当該方言においては一部、起伏型との揺れも見られるものの、Ⅰ類（平板型）の動詞とほぼ同じアクセントになる。次に、各型式のアクセントに着目すると、特にⅠ類（平板型）に特徴が見られる。連用形に「行く」が接続した形式のアクセントで、「アスビꜛイク」のように動詞の末尾に下がり目があるのは県内の他の地点でも観察されている事象である。また、「ダローꜛ」、「ナカッタ」が接続したアクセントは西関東方言におけるアクセントの古相を残している。また、後続形式に下がり目のある「ダンベ」や「ダトヨ」が接続したアクセントも注目すべき点である。

表３：三浦市方言老年層話者　動詞の活用形のアクセント

後続形式 / 語例	終止形 命令形	～タ （た） ～テ （て）	～ナ（な・禁止） ～バ（ば・仮定）	～コト（こと・事） ～イル（れる・受身、尊敬） ～セル（せる・使役） ～テー（たい・願望） ～ネー（ない・打消）	～イク（しに行く）	～ノガ（方が、物が） ～ヤー（ば・仮定） ～ベー（よう・意思・勧誘） ～マス（ます）
遊ぶ （3拍・Ⅰ類）	アスブ アスベ	アスンダ アスンデ	アスブꜛナ アスベꜛバ	アスブコト アスバイル アスバセル アスビテー アスバネー	アスビꜛイク	アスブノꜛガ （アスブꜛノガとも） アスビャꜛー アスブベꜛー アスビマꜛス
※入る （3拍・Ⅲ類）	ヘール ヘーレ	ヘーッタ ヘーッテ	ヘꜛーンナ ヘーレꜛバ	ヘーンコト ヘーライル ヘーラセル ヘーリテー （ヘーリテꜛーとも） ヘーラネー （ヘーラꜛネー、	ヘーリꜛイク	ヘーンノꜛガ （ヘールꜛノガとも） ヘーリャ ヘーンベꜛー ヘーリマꜛス
選ぶ （3拍・Ⅱ類）	エラꜛブ エラꜛベ	エラꜛンダ エラꜛンデ	エラꜛブナ エラꜛベバ	エラꜛブコト エラバイꜛル （エラバꜛイルとも） エラバセꜛル エラビテꜛー エラバꜛネー	エラꜛビイク	エラꜛブノガ エラꜛビャー エラブベꜛー エラビマꜛス
帰る（3拍・Ⅱ類）	ケꜛール ケꜛーレ	ケꜛーッタ ケꜛーッテ	ケꜛーンナ ケꜛーレバ	ケꜛーンコト ケーライꜛル （ケーラꜛイルとも） ケーラセꜛル ケーリテꜛー ケꜛーンネー	（ケꜛーリイク）	ケꜛーンノガ ケꜛーリャ ケーンベꜛー ケーリマꜛス

後続形式 語例	～ソーダ（そうだ・様態） ～ナガラ（ながら）	～ダロー（だろう・推量） ～トヨー（そうだ・伝聞） ～ラシー（らしい）	～ダンベ（だろう） ～ダトヨ（そうだ・伝聞） ～ホーガ（方が） ～ヤガレ（やがれ）	～ナカッタ（なかった） ～ッサンナ（するな・禁止）
遊ぶ （3拍・Ⅰ類）	アスビソーダ アスビナガラ	アスブダロー アスブトヨー アスブラシー アスバッセー アスブベヨー	アスブダンベ アスブダトヨ アスブホーガ アスビヤ゛ガレ （アスビャ゛ガレとも）	アスバナカッタ （アスバナカッタとも） アスバッサンナ
※入る （3拍・Ⅲ類）	ヘーリソーダ ヘーリナガラ	ヘーンダロー ヘーントヨー ヘーンラシー ヘーラッセー ヘーンベヨー	ヘーンダンベ ヘーンダトヨ ヘールホーガ ヘーリヤ゛ガレ （ヘーリャ゛ガレとも）	ヘーンナカッタ （ヘーンナカッタとも） ヘーラッサンナ
選ぶ （3拍・Ⅱ類）	エラビソーダ エラビナガラ	エラブダロー エラブトヨー エラブラシー エラバッセー エラブベヨー	エラブダンベ エラブダトヨ エラブホーガ エラ゛ビヤガレ （エラ゛ビャガレとも）	エラバナカッタ （エラバナカッタとも） エラバッサンナ
帰る（3拍・Ⅱ類）	ケーリソーダ ケーリナガラ	ケーンダロー ケーントヨー ケーンラシー ケーラッセー ケーンベヨー	ケーンダンベ ケーンダトヨ ケーンホーガ ケ゛ーリャガレ	ケーンナカッタ ケーラッサンナ

動詞のもう一つの問題点として、複合動詞のアクセントがある。坂本（2017）では、「オキヤガレ（起きやがれ）」のような、前部要素の下がり目が生かされる例を示したが、複合動詞でも同様に、前部要素の下がり目が保たれるアクセントが観察されることが予想された。そこで、前部要素が起伏型の動詞である複合動詞を中心にアクセントを観察したところ、多くは「アミダシタ・アミダシタ（編み出した）」のように平板型か、末尾の 2 拍前に下がり目のあるアクセントとの間で揺れていたが、「ケッタボス（蹴倒す）」や「オモイシル（思い知る）」、「マチカメール（待ち構える）」など、より古いと考えられるアクセントも観察された。

４．５形容詞のアクセント

伝統相を多く残す名詞、動詞とくらべると、形容詞のアクセントの実態には変化の兆候が観察された。終止形単独のアクセントは、「アセー（浅い）」、「タケー（高い）」のように対立が保たれている語もあるが、同音異義語の「熱い・暑い」と「厚い」の区別が判然としないとか、「アケー・アケー（赤い）」と両方の方で発音される語があるなど、平板型から起伏型への統合が生じているような傾向が見える。中郡二宮町県内で、語尾の連母音が融合すると平板型の形容詞が起伏型で発音されやすくなる地域があるが、当該方言では、「オセー（遅い）」と融合しても平板型が保たれていたり、「カルイ（軽い）」と非融合形でも下がり目が観察されたりと、語尾の連母音の融合は起伏型への統合と関与していない様子であった。活用形のアクセントでは平板型と起伏型の対立は保たれているが、それぞれの形式で揺れを生じている部分がある。たとえば、起伏型の頭高型の形式は、下がり目が 1 拍後退した中高型との間で揺れている。また、「～カッタ」と「～ケレバ」の形式のアクセントの揺れは同一のアクセントの型への変化の方向と言える。

表４：三浦市方言老年層話者　形容詞の活用形のアクセント

後続形式 ＼ 語例	終止形	～ヨ（よ） ～カ（か・疑問） ～ダ（だ・断定） ～テ（て） ～タ（た） ～バ（ば・仮定） ～キャ（ば・仮定）	～コト（事） ～デス（です） ～カラ（から） ～ジェン（ではないか） ～ノガ（方が、物が） ～ッテ（て） ～ネー（ない・打消） ～カネー（～はない） ～ナル（なる）	～ベー（べー・推量）	～ソーダ（そうだ・様態） ～トヨー（そうだ・伝聞）	ラシー（らしい）
浅い （3拍・Ⅰ類）	アセー	アセーヨ アセ┐ーカ アセ┐ーダ アサ┐クテ アサカ┐ッタ （アサ┐カッタと揺れ） アサケ┐レバ （アサ┐ケレバと揺れ） アサ┐キャ	アセーコト アサ┐イデス アセ┐ーカラ アセ┐ージェン アセ┐ーノガ アサ┐クッテ アサクネ┐ー アサ┐カネー アサクナ┐ル	アサカンベ┐ー	アサソーダ アセ┐ートヨー	アセーラシ┐ー
高い （3拍・Ⅱ類）	タケ┐ー	タケ┐ーヨ タケ┐ーカ タケ┐ーダ タ┐カクテ※ タ┐カカッタ※ タ┐カケレバ※ タ┐カキャ※	タケ┐ーコト タカ┐イデス タケ┐ーカラ タケ┐ージェン タケ┐ーノガ タ┐カクッテ※ タ┐カクネー※ タ┐カカネー※ タ┐カクナル※	タカカンベ┐ー	タカソ┐ーダ タケ┐ートヨー	タケーラシ┐ー

なお、表中の※は「タカ┐～」との揺れを表す。

５．結論と今後の課題

三浦市方言のアクセントは東京式アクセントの体系を持つ。名詞と動詞のアクセントからは、当該地域やその周辺の地域のアクセントの伝統相が観察される一方で、形容詞は起伏型への統合や、新しいアクセントの型との間での揺れが見られるなど、変化が目立つ結果となった。

課題は、話者１名に対して調査を行っている点である。観察されたアクセントがどのような性質であるかは、先行研究と共通していればわかるが、合致しないデータについては複数の話者のデータをみて確認する必要がある。調査を継続し、データを充実させたい。

参考文献

秋永一枝（2001）「東京アクセント習得法則」金田一春彦・監『新明解日本語アクセント辞典』三省堂

坂本薫（2017）「神奈川県三浦市方言にみられる音声、音韻の特徴」第 335 回日本音声学会研究例会発表資料

東條操（1954）『日本方言学』吉川弘文館

日野資純（1952）「相模方言の素描（その方言区画）」『国語学』９集　武蔵野書院

日野資純（1964）「神奈川県の方言区画」『日本の方言区画』日本方言研究会編　東京堂

日野資純（1984）「アクセント研究に対する一つの提言（アクセント類別語彙未登載語のアクセントと方言研究）」『音声学会会報』

日野資純（1992）「神奈川県方言」平山輝男ほか編『現代日本語方言大辞典』１巻　明治書院

平山輝男・編（1960）『全国アクセント辞典』東京堂

P12

心的態度を示すドイツ語心態詞と日本語終助詞の発話における韻律的特徴について[1]

生駒　美喜（早稲田大学）
ikoma@waseda.jp

1. はじめに

ドイツ語の日常会話には，心態詞と言われる様々な心的態度を示す語が用いられる．心態詞は，状況によって様々な話者の意図やニュアンスを表す．例えばドイツ語の schon は文脈によって，1)「きっと」という意図の「確信」，2) 相手の言ったことを肯定しつつ一部留保する「留保付肯定」，3) 相手が否定したことに対する「反論」という 3 つの意図を表すことができる(Ikoma 2007)：

(1) a. Peter kommt schon. ペーターは（きっと）来るよ．「確信」
b. Peter kommt schon. ペーターは来るよ．（けど・・・）「留保付肯定」
c. Peter kommt schon. （いや,）ペーターは来るよ．「反論」

Ikoma (2007)では，心態詞 doch, ja, schon, denn を含む発話において，話者の意図やニュアンスによって特有の韻律的特徴が見られることが明らかになった．また，生駒(2014)は,「反論」の意図を示す心態詞 schon，doch を含む発話を分析した．その結果，ピッチアクセントの置かれる動詞部分のピッチピークがより遅く生じることで,「反論」として知覚されることが示唆された．さらに，Ikoma (印刷中)では，上記(1)c.の短文を用いた「反論」の発話の韻律的特徴を，同じ位置にアクセントが置かれる「確信」(1)a.の発話と比較した結果,「反論」の発話は「確信」の発話と比較して発話全体と各音節の母音の持続時間がより長く，発話全体の平均基本周波数(F0)および F0 最大値，最小値が「確信」と比べてより低くなっていた．また，アクセントの置かれる音節開始部分の F0 の値が「確信」と比較してより低く，F0 ピーク値がより高くなることで，アクセントのある音節の F0 の上昇幅がより大きくなっており，アクセント部分の F0 のピークがより遅く生じていることが明らかになった．

一方，ドイツ語心態詞と似た機能を持つとされる日本語終助詞（Kawamori 1997）の発話にも，その意味・機能に韻律的特徴が関わっていることが次第に明らかにされている（片桐 1997, 大島 2013, 郡 2015）．さらには,「感心」や「疑い」などのパラ言語情報において，持続時間，ピッチなどの韻律的特徴が深く関わっていることが以前から指摘されている（森他 2014）．では，共通する心的態度を示すドイツ語心態詞を含む発話と日本語終助詞を含む発話には，それぞれどのような韻律的特徴が見られるのか．同じ心的態度を示す発話に共通した韻律的特徴は見られるだろうか．

本研究では，Ikoma（印刷中）における心態詞 schon を含む発話と状況に対応する日本語の終助詞「よ」を含む発話と状況文を用いて発話実験を実施し，音響分析を行い，日本語

[1] 本研究は, JSPS 科研費 24520441 および早稲田大学特定課題(B) 2017B-004 の助成を受けています.

終助詞を含む発話の音響特徴を明らかにし，Ikoma（印刷中）におけるドイツ語心態詞の発話の音響特徴と比較する．

2. 発話実験

2.1. 実験の資料と被験者

Ikoma（印刷中）の発話実験に用いたドイツ語心態詞 schon を含む短文"Peter kommt schon"の日本語訳となる「ペーターは来るよ」の文を発話文として用い，Ikoma (ebd.)で作成した「確信」「留保付肯定」「反論」それぞれの状況文および対話文を日本語に訳した資料を用意した．実験の被験者として計 10 名（男性 3 名，女性 7 名）の日本語を母語とする大学生に協力してもらった．このうち 8 名は関東地方出身者で，残る 2 名のうち 1 名は北海道出身，もう 1 名は長野県出身である．

2.2. 発話実験の手順

実験は互いに知っている者同士 2 名 1 組の被験者で行った．2 名はそれぞれの録音ブースに入り，互いの声のみが聞こえる状況で録音を行った．被験者は実験資料をよく読んで理解した後，2 名で練習を行った．その後，被験者 2 名は各々の役を演じ，その状況になったつもりで，対話部分のみを 3 回繰り返して朗読[2]した．対話文の役割は途中で交代し，各々が両方の話者の役を演じた．

2.3. 音響分析

録音した音声データ（1 名につき 3 回×3 つの発話意図＝計 9 回の発話文，合計 89 発話[3]）をコンピュータ上に取り込み，Praat (Boersma & Weenink 2017)を用いて音響分析を行い，以下の音響特徴を測定した[4]：

1) 持続時間：発話全体，各音節の持続時間
2) 基本周波数（F0）：発話全体の最大値，最小値，平均値，ピッチレンジ，2 つのアクセント句（AP1:「ペーターは」，AP2:「来るよ」）における最大値，最小値，平均値，ピッチレンジ，AP1，AP2 の句頭の F0 の上昇幅および F0 ピークのタイミング
3) 振幅：発話全体，各音節における振幅の最大値，最小値，平均値，アクセント句（AP1，AP2）における最大値，最小値，平均値，振幅の変動幅

さらに，発話末の「よ」の部分の句末音調（Boundary Pitch Movement, BPM）を調べた[5]．

2.4. 分析結果と考察

2.4.1. 持続時間

図 1 に，「確信」「留保付肯定」「反論」の発話時間における「ペー」「ター」「は」「く」「る」

---

[2] この発話のバリエーションとして，A, B の対話文の前後に間投詞などを自由に入れて読んでもらう発話を同時に録音したが，本研究の分析対象にはしていない．

[3] 1 名の被験者において 2 回しか繰り返されていない発話があったため，89 発話となった．

[4] 音響分析に際し，小西隆之氏（早稲田大学国際教養学部助手）に Praat スクリプト作成にご協力いただいた．また，小西氏とユン・ジヒョン氏（上智大学大学院理工学研究科博士後期課程）にアノテーション作業にご協力いただいた．ここに感謝申し上げる．

[5] 小西隆之氏に聴取とアノテーションを行っていただいた．

「よ」それぞれの音節の持続時間および発話全体の持続時間を示す.

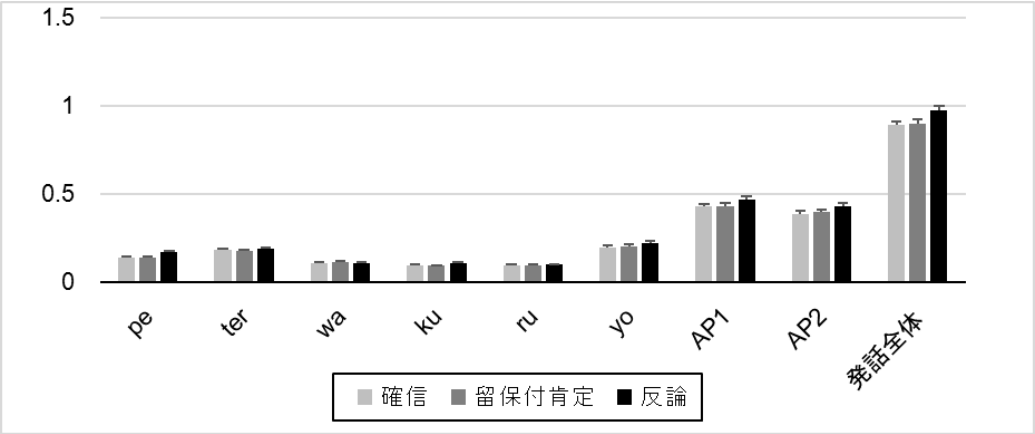


図1：各音節，AP1, AP2，発話全体の持続時間（単位：sec）

発話の最初の音節「ペー」と最後の音節「よ」，AP1，AP2，発話全体の持続時間において，「反論」がその他の発話意図である「確信」および「留保付肯定」よりも長くなっていた[6]. Ikoma （印刷中）におけるドイツ語心態詞を含む発話においては，文全体およびすべての母音部分の持続時間が「反論」において「確信」よりも長くなっていた[7]．発話全体の時間が「反論」の意図を含む発話において長くなっていることは，日本語とドイツ語の発話に共通する特徴として捉えることができるであろう.

2.4.2. 基本周波数(F0)

発話全体およびアクセント句（AP1，AP2）の基本周波数の値を図 2 に示す．F0 の最大値は，発話全体および 2 つのアクセント句において，いずれも「反論」が他の 2 つの発話意図より高いことが明らかになった[8]．また，F0 変動幅（ピッチレンジ）に関しては，「留保付肯定」における発話全体のピッチレンジが他の 2 つの発話意図と比較して小さかった[9]. この結果は，「留保付肯定」を示すドイツ語心態詞の発話の特徴とも共通している[10]．AP1, AP2 のピッチレンジは「反論」が他の 2 つの発話意図よりも大きいことが分かった[11].

---

[6] 一元配置分散分析の結果，有意差が見られた："pe": $F(2, 56) = 12.286$, $MSe=0.001$, $p< .01$, "yo": $F(2, 56) = 7.055$, $MSe=0.001$, $p< .01$, AP1: $F(2, 56) = 5.462$, $MSe=0.003$, $p< .01$, AP2: $F(2, 56) = 8.992$, $MSe=0.002$, $p< .01$, sentence: $F(2, 56) = 8.858$, $MSe=0.007$, $p< .01$

[7] Ikoma においては，アクセントの位置が同じ「確信」と「反論」の発話を主として分析対象としているため，「留保付肯定」に関しては詳しい分析結果が出ていない.

[8] 一元配置分散分析の結果以下の通り有意差が見られた．sentence: $F(2, 56) = 9.597$, $MSe=6.418$, $p< .01$, AP1: $F(2, 56) = 6.792$, $MSe= 8.344$, $p< .01$, AP2: $F(2, 56) = 13.258$, $MSe= 6.953$, $p< .01$

[9] 一元配置分散分析の結果以下の通り有意差が見られた．$F(2, 56) = 7.06$, $MSe=16.47$, $p< .01$

[10] Ikoma（印刷中）では「反論」の発話に着目して分析を行っているため，「留保付肯定」に関するデータは示されていないが，本研究との関連で，Ikoma (ebd.)における音声データを再分析したところ，発話全体におけるピッチレンジが「留保付肯定」において他の発話意図と比べて小さいことが明らかになった.

[11] 一元配置分散分析の結果以下の通り有意差が見られた．AP1: $F(2, 56) = 7.086$, $MSe= 8.305$,

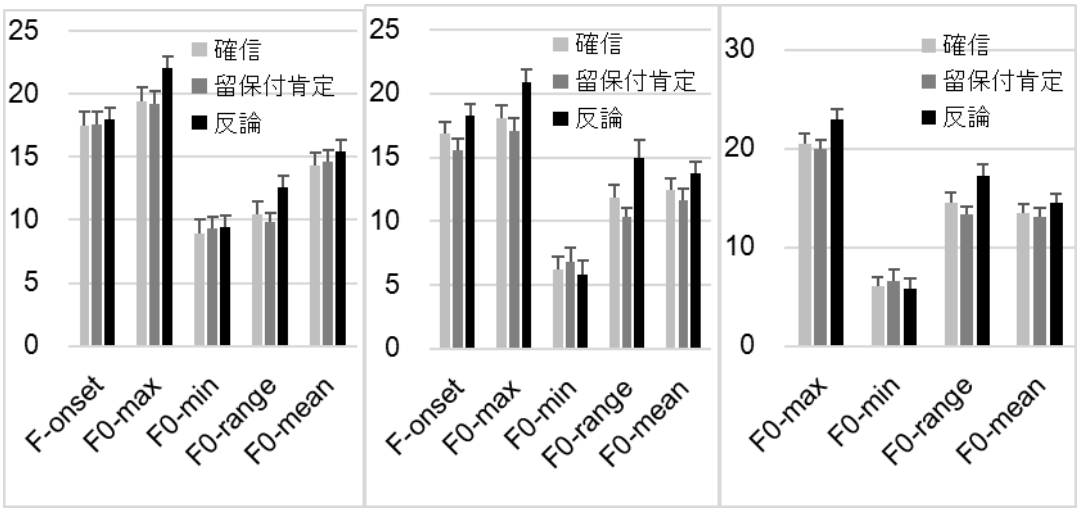


図 2: 発話全体（左）における F0 の最大値，最小値，ピッチレンジ，平均値，およびアクセント句 AP1（中央），アクセント句 AP2（右）における F0 の開始値，最大値，最小値，ピッチレンジ，平均値（単位：semitone）

日本語では，アクセント句の冒頭でのピッチ上昇（「句頭の上昇」）が，発話意図と関わっていることが知られている（森他 2014:82）．そこで，2 つのアクセント句 AP1, AP2 における句頭のピッチの上昇幅および上昇のタイミングを測定した．その結果を図 3 に示す．

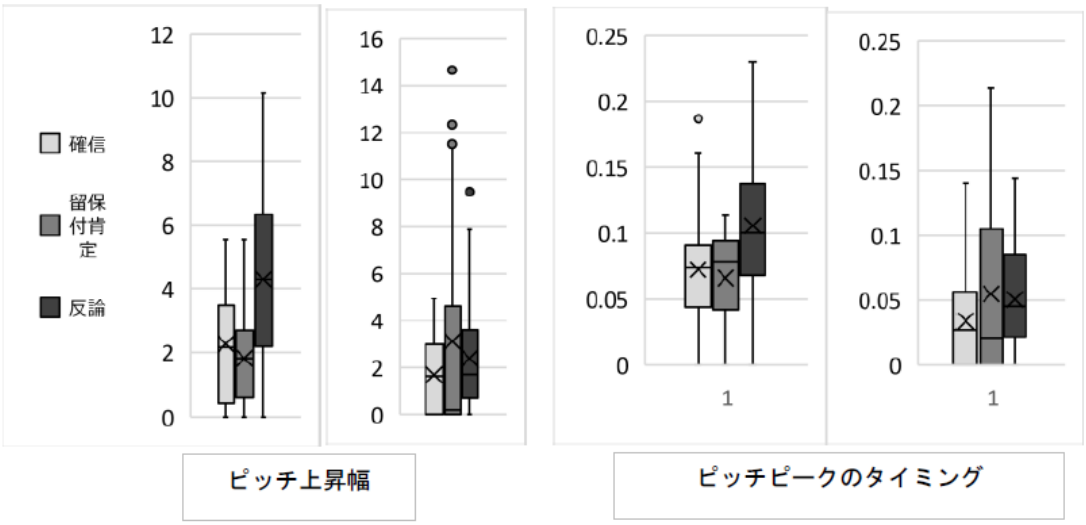


図 3 アクセント句 AP1（左 1 番目），アクセント句 AP2（2 番目）における冒頭のピッチ上昇幅（単位：semitone），アクセント句 AP1（3 番目），アクセント句 AP2（4 番目）におけるピッチピークのタイミング（単位：sec）

AP1「ペーターは」（図 3，一番左）のアクセント句の冒頭のピッチの上昇幅が，「反論」の発話において他の 2 つの発話意図と比較して大きいことが分かる[12]．一方，AP2「来るよ」

$p < .01$, AP2: $F(2, 56) = 9.397$, $MSe = 18.172$, $p < .01$

[12] 一元配置分散分析の結果有意差が見られた：$F(2, 56) = 14.58$, $MSe = 3.29$, $p < .01$

（図 3，左から 2 番目）のピッチの上昇幅においては，3 つの発話意図に有意差は見られなかった[13]．ピッチのピークが生じるタイミングは AP1(図 3，左から 3 番目）の「反論」の発話において他の 2 つの発話意図と比較して遅いことが分かる[14]．一方 AP2（図 3，4 番目）においては，「反論」の発話におけるピッチのピークのタイミングには差異が見られなかった[15]が，AP2 のどの音節に F0 ピークが生じているかを調べた結果，「留保付肯定」では「来るよ」の「く」の 90%に F0 ピークが生じているのに対し，「確信」「反論」の発話の約 60%に F0 ピークが生じ，残る発話では後続の音節「る」に F0 ピークが生じていた．この結果から，発話意図がピッチの下降タイミングにも影響している可能性が示唆される．

2.4.3. 振幅

発話全体および AP1, AP2 における振幅の最大値は，「反論」の意図を示す発話において，他の発話意図と比較して大きいことが明らかになった[16]．

2.4.4. 句末音調

終助詞「よ」の部分にどのような句末音調が現れるか，「確信」「留保付肯定」「反論」それぞれの発話を調べた[17]結果，「留保付肯定」の約 60%に H の音調が現れることが分かった．但し，H の音調には上昇調と平板調とが存在するため，今後の研究にて詳しく調べたい．

3. まとめ

本研究においては，「確信」「留保付肯定」「反論」の意図を示す日本語の終助詞を含む発話「ペーターは来るよ」の音響分析を行い，韻律的特徴を調べ，ドイツ語心態詞の発話の特徴と比較した．その結果，以下の点が明らかになった：

1) 持続時間：発話全体の持続時間が「反論」において長くなっていた．この結果はドイツ語の心態詞を含む「反論」の発話とも共通している．一方ドイツ語心態詞の発話では「反論」の持続時間が全ての音節で伸びているのに対し，日本語終助詞を含む発話では最初と最後の音節のみ持続時間が伸びていた．
2) 基本周波数：発話全体の基本周波数（F0）は「反論」の発話の最大値が他の 2 つの発話意図よりも高かった．これに対し，ドイツ語心態詞の発話では「反論」において F0 最大値が低くなっていた．ドイツ語の発話では全体のピッチを下げることによって「反論」の意図を示すのに対し，日本語の発話では，全体のピッチを上げることで「反論」の意図を示すことが以上の結果から示唆される．発話全体のピッチレンジは，「留保付肯定」が他の 2 つの発話意図と比較して小さかった．ドイツ語の発話においても同様の結果が見られており，今後詳しく分析をする必要がある．さらに日本語の「反論」の発話では，最初のアクセント句の句頭上昇幅が大きく，ピークタイミングが遅く生

---

[13] 一元配置分散分析の結果有意差は見られなかった：$F(2, 56) = 1.102$, $MSe = 16.757$, *n.s.*
[14] 一元配置分散分析の結果有意差が見られた：$F(2, 56) = 8.813$, $MSe = 0.001$, $p < .01$
[15] 一元配置分散分析の結果有意差は見られなかった：$F(2, 56) = 2.077$, $MSe = 0.002$, *n.s.*
[16] 一元配置分散分析の結果有意差が見られた：AP1: $F(2, 56) = 6.408$, $MSe = 6.981$, $p < .01$, AP2: $F(2, 56) = 14.637$, $MSe = 7.564$, $p < .01$, sentence: $F(2, 56) = 10.968$, $MSe = 6.481$, $p < .01$.
[17] 小西氏に句末音調の聴取とアノテーションをしていただいた．

じることが明らかになった．ドイツ語の「反論」の発話においても，アクセントのある音節冒頭の上昇幅が大きく，ピークタイミングがより遅くなるという類似の結果が見られており，今後詳しく分析する必要がある．

3) 振幅：日本語の「反論」を示す発話において，発話全体およびアクセント句の振幅の最大値がより大きいという結果が見られた．ドイツ語心態詞の発話に関しては今後分析を行い，共通点，相違点を明らかにしていきたい．

本研究では，日本語終助詞を含む発話の韻律的特徴を見てきたが，発話意図には韻律的特徴の他に母音の音質や発声様式（森他 2014）などの音響特徴が関わっている．今後も分析を行い，ドイツ語心態詞を含む発話と日本語終助詞を含む発話に共通にみられる特徴，言語固有にみられる特徴を明らかにしていきたい．

参考文献

Boersma, Paul / Weenink, David (2017) *Praat: doing phonetics by computer* [Computer program]. Version 6.0.29, retrieved (15 July 2017) from http://www.praat.org/

Ikoma, Miki (2007) *Prosodische Eigenschaften der deutschen Modalpartikeln*. (Schriftenreihe PHONOLOGIA, Band 103). Hamburg, Dr. Kovač.

Ikoma, Miki（印刷中）"Prosodie und Bedeutung der unbetonten und betonten Modalpartikel *schon*", *Akten des 44. Linguisten-Seminars, Tokyo 2016*. München, Iudicium.

生駒美喜 (2014) 「反論の意図を表すドイツ語心態詞を含む発話の韻律的特徴について」. 『第 28 回日本音声学会全国大会予稿集』日本音声学会, 123-128.

片桐恭弘(1997) 「12. 終助詞とイントネーション」音声文法研究会編『文法と音声』(pp.235-256), くろしお出版.

Kawamori, Masahito (1997) "Epistemic Functions of Japanese Sentence Final Particles" In: Kajita, Masaru et al.（eds.）*Studies in English Linguistics: A Festschrift for Akira Ota.* Tokyo, Taishukan. 大修館, 1002-1015.

郡史郎(2015) 「終助詞『ね』のイントネーション」『音声言語の研究』10（大阪大学大学院言語文化研究科）: 61-76. http://doi.org/10.18910/57286 (2017.7.30.閲覧)

森大毅・前川喜久雄・粕谷英樹 (2014)『音声は何を伝えているか：感情・パラ言語情報・個人性の音声科学』（日本音響学会編音響サイエンスシリーズ 12）コロナ社.

大島ディヴィッド義和(2013)「日本語におけるイントネーション型と終助詞機能の相関について」『国際開発研究フォーラム』43: 47-63. http://www.gsid.nagoya-u.ac.jp/bpub/research/public/forum/43/04.pdf (2017.7.30.閲覧)

P13

Assessing the Pronunciation of Japanese Leaners of English

Chika Fujiyuki(1), Sayoko Eguchi(2,1), Reiko Akahane-Yamada(2,1)
(1) Graduate School of Intercultural Studies, Kobe University
(2) Advanced Telecommunications Research Institute International (ATR)
141c306c@stu.kobe-u.ac.jp, eguchi@atr-lt.jp, yamada@atr-lt.jp,

1. Introduction

There is a trend that native-like pronunciation is neither a realistic nor a suitable goal for leaners of English as a second language (Isaacs, 2013). The absence of pronunciation in rating scales of frequently used speaking tests reflects this fact. Even if pronunciation is included in a scale, it is evaluated mainly in terms of intelligibility or comprehensibility (cf. Appendix 1). Having a good command on segmental and prosodic features, however, will encourage leaners to speak out. The present study attempted to provide detailed pronunciation assessment using multiple metrics. Utterances produced by native speakers of Japanese on an English word reading-aloud task were evaluated by a phonetically trained Japanese-English bilingual. The correlation among the metrics was also examined to explore what property contributes to good pronunciation.

2. Method

2.1. Utterances

Utterances obtained in Eguchi and Yamada (manuscript under preparation) were used for the study. Among them, twenty English words produced by ten native speakers of Japanese on a word reading-aloud task were used. The age of the speakers, including both males and females, ranged from 19 to 40. Their English ability in terms of grade or score on English tests varied from EIKEN Grade 3 to TOEIC score 940, and none of them has been lived or studied abroad more than 12 months. The words used were following twenty words of low and high familiarity:

ahead	allegory	amass	aperitif	badminton
belfry	believe	bring	calcium	caul
Christmas	coate	conciliatory	curriculum	delicious
derelict	diode	disavow	discriminatory	doting

2.2. Assessment

Assessment was done by a phonetically trained Japanese-English bilingual. Each utterance was first transcribed with IPA and then evaluated using six metrics: 1. phoneme substitution; 2. elision; 3. epenthesis; 4. primary stress; 5. rhythm; 6. holistic goodness. One-syllable words were excluded in the evaluation with metric 4 and 5. Occurrence of phoneme substitution, elision and epenthesis was counted respectively for each speaker. Primary stress was evaluated by being classified into three

groups : no stress placement, stress placement on a wrong syllable, and stress placement on a right syllable. The ratio of the count in the third category to the entire count was obtained as accuracy in primary stress, and accuracy was obtained for each speaker and for each word. Metric 5, rhythm, was rated on a scale of 1 to 3 (1=not good, 3=good) and metric 6, holistic goodness, of 1 to 5 (1=not good, 5=good). Average scores for metric 5 and 6 were calculated by speaker and by word.

3. Result

3.1. Phoneme substitution, elision, epenthesis

Frequent occurrences were:

- Substitution of weak vowel /ə/ with /a/, /ʌ/ or /eɪ/
- Substitution of /ɪ/ with /i/
- Substitution of /aʊ/ with /oʊ/
- Substitution of /ɹ/ with /ɾ/ (alveolar flap)
- Elision of /oʊ/ (as in -tory)

A strong correlation was found between holistic goodness and substitution (r = 0.95, Fig.1), but not between holistic goodness and epenthesis (r = 0.63, Fig.2) or elision (r = 0.26, Fig.3).

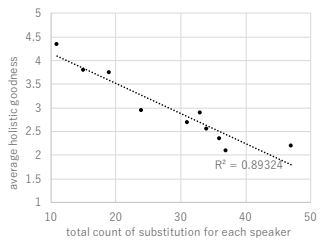


Fig.1 Correlation between substitution and holistic goodness

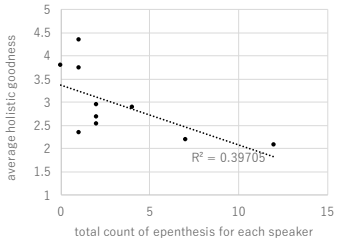


Fig.2 Correlation between epenthesis and holistic goodness

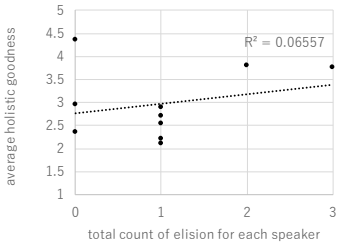


Fig.3 Correlation between elision and holistic goodness

3.2. Primary stress

Primary stress was perceived in all the bi- and polysyllabic words, except *coate*, which was pronounced as monosyllabic by most of the speakers. The accuracy by speaker varied widely, from 41% to 94 %. A strong correlation was evident between holistic goodness and the average by speaker (r = 0.86, Fig.4) as well as the average by word (r = 0.79, Fig.5).

3.3. Rhythm

The average scores by speaker varied from 1.41 to 2.76 and the average scores by word varied from 1.0 to 2.8. A very strong correlation was found between holistic goodness and the average rhythm by speaker (r = 0.97, Fig. 6) as well as the average rhythm by word (r = 0.95, Fig.7).

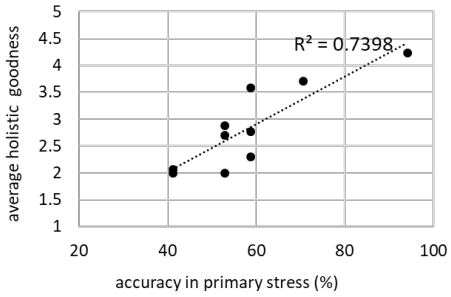


Fig.4 Correlation between holistic goodness and primary stress (by speaker)

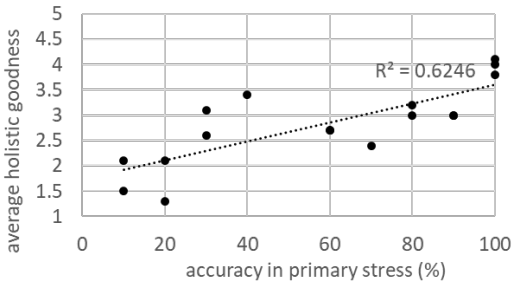


Fig.5 Correlation between holistic goodness and primary stress (by word)

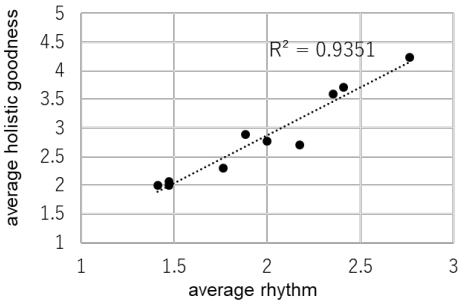


Fig.6 Correlation between holistic goodness and rhythm (by speaker)

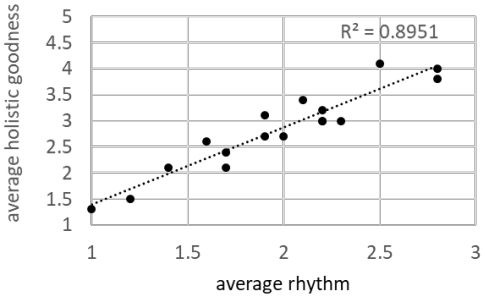


Fig.7 Correlation between holistic goodness and rhythm (by word)

3.4. The number of syllable and holistic goodness

The result of one-way ANOVA [F (4, 45) = 2.578, $p < 0.01$] suggested that the average holistic goodness varied significantly by the number of syllable. A post hoc pairwise comparisons showed that the average holistic goodness in the one-syllable group was significantly higher than the other syllable groups while the average in the six-syllable group was significantly lower than the other groups. The average of two-, three- and four-syllable groups were mutually comparable (Fig. 8).

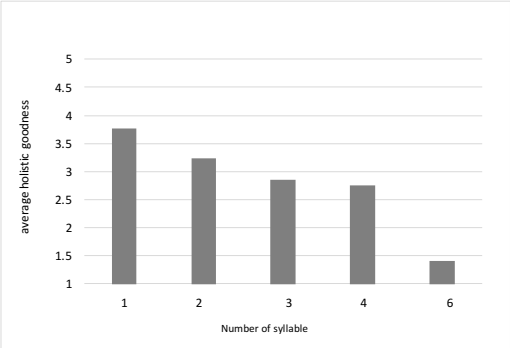


Fig.8 Correlation between the number of syllable and holistic goodness

4. Discussion

Apart from a well known substitution of / ɹ/ with /ɾ/ by Japanese speakers, the substitution of a weak vowel /ə/with /a/ or /ʌ/ as well as /ɪ/ with /i/ were highlighted in this study. This suggests producing weak/lax vowels is a hurdle for many Japanese leaners. Since weak vowels are an essential component of English rhythm, leaners must overcome this barrier to improve their pronunciation. The correlation analysis indicates a strong association between holistic goodness and phoneme substitution, primary stress and rhythm. This result indicates improving both segmental and supersegmental features is requisite for good pronunciation. Evaluating more utterances by multiple raters is needed to enhance the reliability of the findings. Some of the metrics used in this study requires phonological expertise, yet providing detailed pronunciation assessment is manageable: it took for the rater about three to four minutes to evaluate a test word. Accumulating detailed pronunciation assessment is beneficial not only to deeper and broader understanding of the Japanese accent but also to the development of pronunciation instruction material. Such assessment data can also be applied to computer-based pronunciation assessment.

References

Eguchi, S. and Akahane-Yamada, R. (unpublished)

Isaacs, T. (2014). "Assessing pronunciation." In A. J. Kunnan (Ed.), *The companion to language assessment*. (pp. 140-155). Hoboken, NJ: Wiley- Blackwell

Appendix.1: Extracts from speaking skills rating in frequently used English tests

TEST	Criteria	Rating	DESCRIPTION
TOEIC	• Pronunciation (when reading aloud, pronunciation is:)	High Middle Low	Highly intelligible Generally intelligible Not intelligible
	• Intonation and stress levels (refer to the ability to use emphases, pauses, falling-rising pitch to convey meaning)	High Middle Low	Highly effective Generally effective Generally not effective
TOEFL	• Delivery difficulties with pronunciation or intonation patterns that effect intelligibility • Difficulties that requires listeners effort	4 3 2 1	Minor difficulties which do not affect overall intelligibility Minor difficulties which may require effort at times but do not affect intelligibility significantly Basically intelligible, though listener effort is needed because of unclear articulation, awkward intonation, or choppy rhythm and pace Consistent difficulties that cause considerable listener effort
CEFR (Common European	Pronunciation and intonation are omitted		

Framework of Reference for Languages)			
IELTS (International English Language Testing System) *public version	Pronunciation	Band.9 (highest)	• Use a full range of pronunciation with precision and subtlety • Sustains flexible use of features throughout • Is effortless to understand
		Band 6	• Use a range of pronunciation features with mixed control • Shows some effective use of features but this is not sustained • Can generally be understood throughout, though mispronunciation of individual words or sounds reduced clarity at times
		Band 4	• Use a limited range of pronunciation features • Attempts to control features but lapses are frequent • Mispronunciation are frequent and cause some difficulty for the listeners
		Band 3	Shows some of the features of Band 2 and some, but not all, of the positive features of Band 4
		Band 2 (lowest)	Speech is often unintelligible

P14

英語母語話者と日本人英語学習者の冠詞の音声特徴に関する一考察

米山 聖子（大東文化大学）・山根 典子（広島大学）・藤森 敦之（静岡大学）・
吉村 紀子（静岡県立大学）
yoneyama@ic.daito.ac.jp, yamanen@hiroshima-u.ac.jp,
fujimori.atsushi@shizuoka.ac.jp, yoshimun@u-shizuoka-ken.ac.jp

1. はじめに

英語はストレス言語であるために，モーラ言語を母語とする日本人英語学習者にとっては獲得が難しいことは広く知られているところである．ストレスは英語の音韻的特徴に深く関与しており，外国語学習者にとってはとても悩ましいものである（北原 2016）．英語の強音節と弱音節が交互に起こる独特のストレスパターンは，英語の語彙とも深く関わりを持っている．英語の語彙は名詞，動詞，形容詞などの意味理解に重要な役割を果たすと考えられる内容語と冠詞，代名詞，前置詞などの文法機能に関連のある機能語とに分けられる．前者には語強勢を受ける強音節が存在し，強音節は一般的にピッチを高く，発話長を長く，そして強く発話される．これに対し，後者は通常ニュートラルなコンテクストでは弱音節として現れ，しばしば弱化の対象となる．ピッチは低く，発話長は短く，そして弱く発話される．日本人英語学習者の英語発話においては，母語である日本語の音韻特性が影響しており，語彙アクセント，音節をまたいだ時間的な構造や等時性に関わっていることが報告されている（e.g., Kubozono, 1989; Ueyama, 2000; Yamane, Yoshimura, and Fujimori, 2016a, 2016b）．

Yamane, Yoshimura and Fujimori (2016b)では，機能語である代名詞 her の音響的特性と文法構造について検討した．代名詞 her が動詞の目的語として再現される条件（e g., saw her）と所有格として再現される条件（e.g., her class）において，her の発話長が句全体の発話長に占める割合を算出し，北米英語母語話者と日本語を母語とする英語学習者のデータを比較した．その結果，her が現れる統語的位置に関わらず，北米英語母語話者の発話における her の割合が日本人英語学習者のものよりも有意に低いことが明らかになっている．また，代名詞 her は同時に高いピッチで発話されていることも明らかになっている．

本研究では，Yamane, Yoshimura and Fujimori (2016b)の結果に基づき，機能語の一つである冠詞を日本人英語学習者がどのように英語名詞句の中で発話するのかについて検討するものである．もし，Yamane らの主張が正しいとするのであれば，冠詞（a, the）は機能語であるために，Yamane らの研究の名詞 her が動詞の目的語として再現される条件（e g., saw her）と所有格として再現される条件（e.g., her class）と同様に，北米英語母語話者の発話における her の割合が日本人英語学習者のものよりも有意に低く，そして高いピッチで再現される可能性が高いと考えられる．本研究では，冠詞（a, the）の発話長とピッチに関する分析を行うことで，仮説の検証を行うものである．

2. 実験

2.1. 被験者

英語初級レベルの日本語を母語とする大学生男女 11 名（日本人英語学習者）と北米英語を母語とする成人男女 10 名（北米英語母語話者）である．

2.2. 録音文

録音されたトークンは，中学校２年生用の英語検定教科書（板垣, 2011）から抜粋した 7 文 58 語から構成されるパラグラフ(物語)である．以下にそのパラグラフを示す．

> *Some years ago, Mr. Sato had a very kind student in his class. She had a pretty name, Aika. Her classmates liked her very much. Sometimes Mr. Sato saw her at school early in the morning. In her hands, she always had very pretty flowers. She picked them from her garden. Everyone in her class loved the colorful flowers.*

録音は静かな教室で一人ずつ行った．A4 1 ページに書かれたパラグラフを一週間前に渡し，練習するように伝えた．録音は普段通りの音読をするように伝えた．録音中に発話間違いが生じ場合には，自分のペースで言い直すように指示した．

2.3. 分析

今回は a very kind student (Condition 1)，a pretty name (Condition 2)，the morning (Condition 3)，the colorful flowers (Condition 4)の 4 つの名詞句に焦点を当てた．測定については，発話長とピッチについて行った．発話長分析については，冠詞の発話長と句全体の発話長を測定し，冠詞の発話長の句全体の発話長に占める割合を算出した．ピッチの分析については，冠詞のピッチの平均値と後続する強音節のピッチの平均値を測定し，後者から前者を引いた値を算出した．

3. 分析結果

3.1. 冠詞の発話長に関する分析

冠詞の発話長の句全体の発話長に占める割合について，一被験者内要因一被験者間要因の混合検定を行った．被験者内要因は Condition の４条件（Condition 1: a very kind student, Condition 2: a pretty name, Condition 3: the morning, Condition 4: the colorful flowers）であり，被験者間要因は Group の２条件（北米英語母語話者，日本人英語学習者）である．従属変数は冠詞の発話長が句全体の発話長に占める割合である．図 1 は 4 条件における冠詞の発話長が句全体の発話長に占める割合をグループごとにプロットしたものである．

Condition に主効果が認められた（Condition: $F(3, 57) = 3.665$, $p = 0.017$）. Condition についての多重比較（Bonferroni 検定）から，Condition 1 (a very kind student)の条件（0.078，SE = 0.007）よりも Condition 4 (the colorful flowers)の条件（0.107, SE = 0.008）のほうが，有意に冠詞の発話長の句全体の発話長に対する割合が高いことが明らかになった（$p = 0.05$）. Group にも主効果が見られた（$F(1,19) = 16.859$, $p < 0.001$）．冠詞の発話長の句全体の発話長に対する割合は北米英語母語話者（0.074, SE = 0.007）のほうが日本人英語学習者（0.116, SE = 0.007）

よりも有意に低かった．Condition と Group の交互作用は認められなかった．

Condition ごとの Group の比較では，Condition 4 (the colorful flowers)以外の 3 条件において北米英語母語話者の句全体の冠詞の割合が日本人英語学習者のものよりも有意に低いことが明らかになった(Condition1 : $F(1,19) = 5.671, p = 0.028$; Condition2 : $F(1,19) = 14.980, p = 0.001$; Condition3 : $F(1,19) = 13.057, p = 0.002$). また，Group ごとの Condition の多重比較（Bonferroni 検定）では，有意差が存在したのは日本人英語学習者の Condition 1 (a very kind student)と Condition 2 (a pretty name)の比較であった．Condition 2 (a pretty name)のほうが Condition 1 (a very kind student)よりも有意に句全体の冠詞の割合が高いことが明らかになった($p = 0.05$).

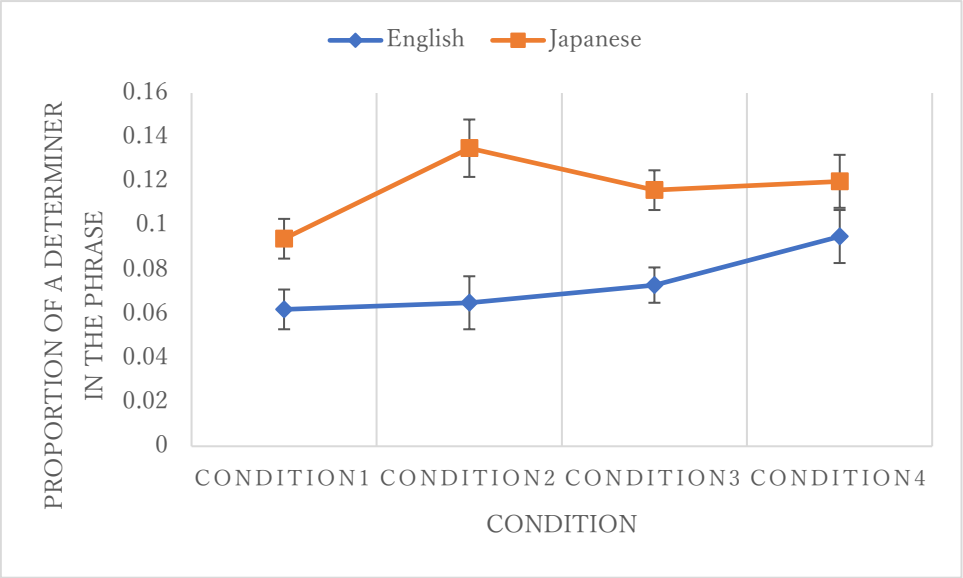


図 1: グループ別4条件における冠詞の発話長が句全体の発話長に占める割合

3.2. 冠詞と後続強音節のピッチに関する分析

冠詞のピッチの平均値と後続強音節のピッチの平均値の差（後続強音節のピッチの平均値から冠詞のピッチの平均値を引いた値）について，一被験者内要因一被験者間要因の混合検定を行った．被験者内要因は Condition の 4 条件（Condition 1: a very kind student, Condition 2: a pretty name, Condition 3: the morning, Condition 4: the colorful flowers）であり，被験者間要因は Group の 2 条件（北米英語母語話者，日本人英語学習者）である．従属変数は冠詞のピッチの平均値と後続強音節のピッチの平均値の差である． 4 つの条件すべての数値が測定できた被験者は北米英語母語話者が 7 名，日本人英語学習者が 8 名であった．これらの被験者データを今回の分析の対象とした．図 2 は 4 条件におけるにおける冠詞と後続強音節のピッチの平均値の差をグループごとにプロットしたものである．

Condition に主効果が認められなかった（$F(3,39) = 1.334, p > 0.05$）． Group には主効果が見られた（$F(1,13) = 12.508, p < 0.005$）．冠詞と後続強音節のピッチの平均値の差の数値は北

米英語母語話者のほうが（-5.819, SE = 3.184）日本人英語学習者よりも（9.602, SE = 2.979）有意に低かった．Condition と Group の交互作用は認められなかった．

Group ごとの Condition の多重比較（Bonferroni 検定）では，北米英語母語話者の Condition 1 と Condition 4 の比較が有意であった（Condition 1: 7.837, SE = 4.432; Condition 4: -24.999, SE = 4.712; $p = 0.001$）Condition ごとの Group の比較では，Condition 4 (the colorful flowers)において北米英語母語話者の冠詞と後続強音節のピッチの平均値の差の数値が日本人英語学習者のものよりも有意に低かった（北米英語母語話者：-24.999, SE = 4.712; 日本人英語学習者：9.052, SE = 4.408; $F(1,13) = 27.849$, $p < 0.001$）．

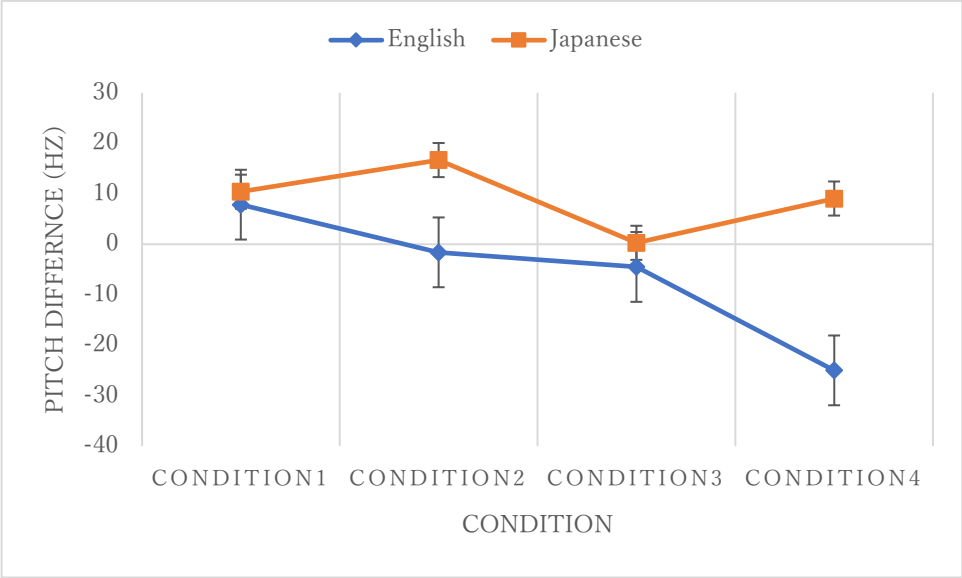


図 2: グループ別4条件における冠詞と後続強音節のピッチの平均値の差

4. 考察

まず，日本人英語学習者の冠詞の発話に関して発話長とピッチ分析の結果から検討を試みる．北米英語母語話者の発話における冠詞の時間長の句全体の時間長に対する割合が日本人英語学習者のものよりも有意に低いことが Condition 4 (the colorful flowers) 以外の Condition 1 (a very kind student)，Condition 2 (a pretty name)，Condition 3 (the morning)の条件について明らかになった．この結果は Yamane ら(2016b)の機能語の一分類である代名詞の場合と同様に，日本語を母語とする英語学習者は，機能語である冠詞を短く発話することができないことを示すものと考えられる．

冠詞と後続強音節のピッチの平均値の差の分析では，日本人英語学習者は全ての条件でプラスの数値を表しており，一貫して冠詞のその直後の強音節のピッチの平均値のほうが冠詞よりのピッチの平均値よりも高い．日本人英語学習者の冠詞と直後の強音節のピッチの比較については，予想通りの傾向を示すといってよいかもしれない．しかしながら，日本

人英語学習者は機能語である冠詞を内容語と同等の扱いをしている可能性も完全には否定できない点と北米英語母語話者のデータとは異なる点を鑑みると，日本人英語話者のピッチデータから冠詞と統語構造を明らかにする決定的な音響特徴ということは現段階では難しいと考えられる．

今回のピッチの分析結果は，日本人英語学習者は名詞句というローカルなドメインでのピッチコントロールについて検討するために行われたが，北米英語母語話者のデータはパラグラフというグローバルなドメインでの音響的なコントロールを示すものとして扱うことができ，日本人英語学習者とは異なるパターンを示している．

北米英語母語話者のデータではパラグラフの最初の文に出てくる Condition 1 (a very kind student)とパラグラフの最後の文に出てくる Condition 4 (the colorful flowers)とでは，２つの分析において有意な違いが見られた．冠詞の発話長の句全体の発話長に占める割合の分析では，Condition 1 (a very kind student)から Condition 4 (the colorful flowers)に向かって割合は増加する傾向にあり，結果的に Condition 1 (a very kind student)と Condition 4 (the colorful flowers)とで有意差が認められた．これに対し，冠詞と後続強音節のピッチの平均値の差の分析では，Condition 1 (a very kind student)から Condition 4 (the colorful flowers)に向かって差の値はプラス値からマイナス値へ変化する傾向にあり，結果的に Condition 1 (a very kind student)と Condition 4 (the colorful flowers)とでは有意差が認められた．

これら二つの現象は別々の現象と考えるのではなく，北米英語母語話者がパラグラフ全体をドメインとした情報構造伝達のプランニングに基づいた時間長とピッチのコントロールの産物であると考えられる．今回の北米英語母語話者のピッチの分析結果は，英語のイントネーション句末でみられる final lengthening と final lowering によって句の時間長の増加とピッチの下降が得られた結果である．Condition 4 (the colorful flowers)はイントネーション句末であると同時にパラグラフの最後の句でもあったため，この側面が強調され，ほかの３条件と明らかに異なる振る舞いが見られたのではないかと推測される．

本研究で明らかにできなかった点も多いが，今後の課題としてここでは３つ挙げておきたい．一つ目は冠詞の発話長に関する分析で日本人英語学習者の Condition 2 (a pretty name)のほうが Condition 1 (a very kind student)よりも有意に句全体の冠詞の割合が高いことが明らかになった点についての解釈についである．確かに Condition 2 は She had a pretty name, Aika. というセンテンスに生起し，Aika という同格に当たる名詞を後ろに持つ点でほかの３条件と異なるが，なぜ Condition 1 (a very kind student)とだけ有意差が生じたかについては説明ができていない．今後，様々なタイプの名詞句を扱うことでこの点については検討していきたい．

２つ目は北米英語母語話者のピッチのデータについてである．ピッチのコントロールは英語においてもグローバルなドメインのイントネーションについてと，ローカルなドメインの音節強勢の両方に関わっているが，今回の北米英語母語話者のデータは前者を示すものであり，後者を示すものではなかった．なぜこのような結果になったのか，またはなぜ前

者と後者の両方を示す結果でなかったのかについて検討していきたい.

最後に，今後は被験者数を増やし，強勢の有無に関係があるとされるラウドネスについての分析も加えて，ピッチ，発話長，ラウドネスの３つの観点から総合的に冠詞の音響的特性と統語構造との関係について分析を試みたい.

参考文献

板垣信哉 (2011) *Sunshine English Course 2*. 東京：開隆堂.

北原真冬 (2016) 「英語のストレスに立ち向かう日本語話者」日本音韻論学会（編）『現代音韻論の動向― 日本音韻論学会 20 周年記念論文集 ―』（pp. 124-127）東京：開拓社.

Kubozono, Haruo (1989) "Syntactic and rhythmic effects on downstep in Japanese", *Phonology*, 6(1), 39-67.

Ueyama, Motoko (2000) *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese*. Doctoral dissertation, University of California Los Angeles.

Yamane, Noriko, Yoshimura, Noriko, and Fujimori, Atsushi. (2016a). "Prosodic transfer from Japanese to English: Pitch in focus marking", *Phonological Studies* 19, 97-104. Tokyo: Kaitakusha.

Yamane, Noriko, Yoshimura, Noriko and Fujimori, Atsushi (2016b) "Japanese EFL Learners' Production of Pronouns in English: Evidence for L2 Prosodic Structure", a paper presented at Acoustics Week in Canada 2016, September 21-23, 2016, The Sutton Place Hotel, Vancouver, British Columbia.

ワークショップ（1） 趣旨説明

プロソディ研究のための方法論：コーパス・生理・文タイプ*

企画　松浦 年男（北星学園大学）

https://researchmap.jp/yearman/PSJ31WS-Prosody/

アクセントやイントネーションなどといったプロソディの研究は，対象とした話者の内省・直観に基づく研究が多く行われている。また，それらの研究と同時並行的に実験的アプローチによる研究も盛んに行われてきた。近年，実験心理学的手法の理解，コーパスの整備・洗練化が進み，プロソディ研究に対するアプローチは広がりを見せつつある。ただし，これらの手法を用いた研究は専用の機材や特にソフトウェアや言語資源に関する知識を要することもあり，敷居の高いものとなりがちであった。しかし，これらを用いることにより，プロソディ一般に対する理解は深まり，また，研究の射程もプロソディだけにとどまらず，言語構造間のインターフェイスや，音声と関連する認知能力などへの応用も期待できる。

そこで，本ワークショップでは発表者たちが行ってきた様々なアプローチによるプロソディ研究について，方法論とその研究手法によって得られた成果に焦点を当てて紹介する。まず，プロソディ研究の主要なテーマとして，アクセントや強制の付与に関するものがある。これらはある単語（無意味語も含む）がどのように発話されるという面に関心が集まっており，産出の側面に焦点を当てたものと言える。その一方で，知覚的側面に関する研究もある。**発表 1** はこのうちアクセントが規則による型から外れたパターンで発話された場合における認知処理について検討する。プロソディの認知的側面がいかに複雑かを知ることができるだろう。

プロソディ研究は様々な単位を対象としており，言語間で差はあるもののおおよそ「発話＞音韻句＞音韻語＞韻脚(フット)＞音節＞モーラ」という階層が仮定されている。**発表 2** はこのうち「音韻句」の形成，特に統語構造とのインターフェイスに焦点を当てたものとなっている。様々な音声的実現がプロソディと統語構造との間のインターフェイスとどのように関わるのか，多様性と普遍性を垣間見ることができるだろう。

さらに，分節音レベルの研究も上記の階層や強勢といったプロソディが関与することが多い。**発表 3** は英語の弾音化を例に，日本語を第一言語とする学習者の発話研究を行うための研究デザインについて，特にコーパスの有用性・必要性に焦点を当てて紹介する。この発表からより精密な研究を行うための手法を学ぶことができるだろう。

このように本ワークショップは広い射程を持っていると同時に，様々な論点を提供するものとなっている。ワークショップでの議論を通じて，方法論に関する理解を深めると同時に，より具体的な問題の提起につなげていくことができれば幸いである。そのため，ワークショップではフロアと今後期待される研究の方向性についても議論し，問題を共有できることを願っている。なお，ワークショップのためのウェブページを用意しているので参照いただけると幸いである。

*本ワークショップは国立国語研究所共同研究プロジェクト「対照言語学の観点から見た日本語の音声と文法」によるものである。

WS1-1 ワークショップ（1）

ERP を用いた複合語アクセントの研究：現状と課題*

松浦 年男（北星学園大学） 安永 大地（金沢大学） 水本 豪（熊本保健科学大学）
yearman@kyudai.jp

1. 序論

本発表ではアクセントのうち規則的な側面に着目し，その知覚における生理的反応について行った実験結果を報告する。実験的手法を用いたアクセントの研究については杉藤美代子氏らによる一連の先駆的な研究が存在し，音響分析だけでなく喉頭筋電図を用いた生理学的特徴や知覚的側面の研究まで網羅している（杉藤(1994-1999)を参照）。一方で近年，事象関連電位（Event-Related Potentials，以下 ERP）を用いた研究も行われてきている。

ERP とはある事象に関連した脳の電位変化のことである。電位成分は ERP 成分と呼ばれ，潜時（頂点潜時），頭皮上分布，実験操作への応答性，極性，波の形状といった点から特定される。これまで言語に関する ERP 成分としては次のものがある（安永(2017/印刷中)を参照）。まず，意味的な逸脱（違反）に対して潜時 400 ミリ秒近辺に観察される陰性成分として N400 がある。単語や文に関する違反に対して潜時 300 から 500 ミリ秒に観察される陰性成分として LAN（Left Anterior Negativity）がある。規則の違反からの修正や修復に対して潜時 600 ミリ秒に観察される陽性成分として P600 がある。

ERP の観察を行った音声研究としては Koso and Hagiwara (2007), Tamaoka ら (2014), 小林ら(2015)などがある。Koso and Hagiwara (2007)では動詞のアクセント違反に対し潜時 400-800 ミリ秒前後に陰性成分が観察され，PAN (Pitch Accent Negativity)と名付けている。一方 Tamaoka ら(2014)はアクセントに特化した ERP 成分の存在を否定している。

日本語のアクセントには語彙的な側面と規則的な側面がある。規則的な側面に関しては，これまで多くの一般化や理論的な説明が行われている（一般化については Kawahara 2015 などを参照）。このうち本稿で焦点を当てる複合語アクセントに関する規則は大きく 2 種類ある。まず 1 つは小林ら(2015)が対象とした韻律語形成に関する規則である。これは「複合語の前部要素と後部要素で 1 つの韻律語を形成せよ」というもので，適用されることにより音韻レベルで 1 語と認定される。もう 1 つはアクセントの型に関する規則である。日本語の複合語（複合名詞）は後部要素によって決まる面が非常に強く（McCawley 1968），小林ら(2015)および本研究が扱う後部要素が 3 モーラの複合語のアクセントついては，次のように一般化できる。

(1) 複合語アクセント規則
   a. 後部要素のアクセントが平板型または最終モーラ→後部要素の第 1 モーラ
      例：おんな]＋ともだち=→女と]もだち，だめ]＋おとこ]→だめお]とこ

* 本研究は国立国語研究所共同研究プロジェクト「対照言語学の観点から見た日本語の音声と文法」，JSPS 科研費（26244022，17H02332，17K02689），平成 28 年度熊本保健科学大学 P&P プロジェクトによるものである。

b. 後部要素が第 1，第 2 モーラ→アクセントを保持（ただし例外あり）
例：な]ま＋たま]ご→生タマ]ゴ（〜タ]マゴも可）

小林ら(2015)は複合語アクセント規則のうち韻律語形成に関する規則に違反したときに PAN 及び P600 が観察されることを指摘している。しかし，複合語アクセントの位置に関する規則(1)については検討を行っていない。韻律語形成も複合語のアクセントに関する規則ではあるが，句レベルの規則でもある。したがって，小林らの結果が本当に「複合語」のアクセント規則を反映したものかは検討を要する。そこで，本研究では(1)の規則に関する実験を実施した。

2. 方法

2.1. 実験参加者

実験は視聴覚に異常がなく，金沢市内に住む大学生 7 名に対し実施した（全員右利きで，神経的な疾患はない）。なおその際，参加者の方言に関してはすべて統制し，東京式あるいは準東京式アクセント体系を有する方言の話者とした。

2.2. 刺激語（調査語）

刺激語として前部要素，後部要素ともに 3 モーラからなる複合語を使用した。後部要素の単語はアクセントに基づき 2 条件に分けた。第 1 の条件は単独発話時に平板型になるもので「平板型条件」と呼ぶ。第 2 の条件は単独発話時に頭高型になるもので「頭高型条件」と呼ぶ。また後部要素の単語はアクセントの他に，親密度による影響を避けるために『日本語の語彙特性』データベースによる親密度が 5.0 以上とし，脳波の計測開始点を一意に定めるために語頭が CVCV ないしは CVN であるものに限定した。それぞれの例を(2)に示す。

(2) 平板型条件：サハラ砂漠，中華ちまき，輸入マグロ…
頭高型条件：ゲーム画面，作業ズボン，笑い涙…

刺激語は第一発表者（1977 年生まれ，東京都出身）によって，正しいアクセント型（複合時に後部要素の第 1 モーラで下がる）と誤ったアクセント型（複合時に平板型になる）で録音された。また，平板型は起伏型に比べ F0 の幅が狭くなるため（Pierrehumbert and Beckman 1988, Kubozono 1988 など），Praat をもちいて F0 の PSOLA 再合成を行い，F0 の幅は平板型に合わせ，平板型もそのままにせず曲線を直線にした。以上により，比較対象となるのは(i)平板型・正常アクセント条件，(ii)平板型・違反アクセント条件，(iii)頭高型・正常アクセント条件，(iv)頭高型・違反アクセント条件という 4 つとなる。

(3) 条件のまとめ

	平板・正常	平板・逸脱	頭高・正常	頭高・逸脱
単独	マグロ=		ズ]ボン	
複合	輸入マ]グロ	輸入マグロ=	作業ズ]ボン	作業ズボン=

2.3. 手続き

脳波の計測には TEAC Polymate AP1532 を用い，電極帽子を通して頭皮上の 19 箇所から脳波を記録した（サンプリング周波数 1000Hz；分析時に 200Hz にダウンサンプリング）。その後，60Hz のノッチフィルター，0.1-30Hz のバンドパスフィルターをかけた信号を加算平均し，条件ごとの ERP を算出した。

3. 結果と考察

3.1. 結果

(i)平板型・正常ア条件と(ii)平板型・違反ア条件を比較すると，(ii)において潜時 700 ミリ秒前後に陰性波が観察され，分散分析によって(i)と比較した場合にその差は有意であった（$F(1, 6)=11.01, p<0.01$）。一方，(iii)頭高型・正常ア条件と(iv)頭高型・違反ア条件を比較すると，(iv)において潜時 300 ミリ秒に陰性波が観察されたが，分散分析によって(iii)と比較した場合にその差は有意ではなかった（$F(1, 6)=0.98, p=0.36$）。図 1 に ERP 波形を示す。実線が正常ア条件，点線が違反ア条件である。

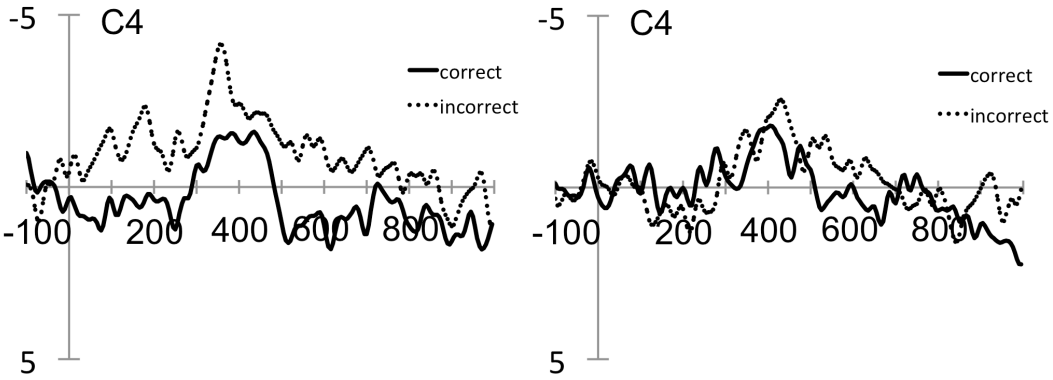


図 1　(i)と(ii)の ERP 波形(左)と(iii)と(iv)の ERP 波形(右)

3.2. 考察

実験の結果は後部要素のアクセントの指定により異なるものであった。(ii)の第 2 モーラにおいて観察された陰性波成分はその特徴から PAN と推定される。すなわち，アクセントに関する逸脱を反映したものだと考えられる。ただし，この成分がアクセントのどのような側面を反映したのかについてはさらなる検討が必要であろう。というのも，単純にアクセントに関する逸脱ということであれば(iv)においても PAN が観察されることが期待されるのに対してそれが観察されていないからである。

(ii)と(iv)の条件間で異なるのは規則適用の余剰性である。(ii)は後部要素が単独時に平板型であるため，複合語アクセント規則が適用されて後部要素の第 1 モーラにアクセントが指定された。それに対して(iv)では，後部要素は単独時に頭高型であるため，複合語の後部要素となったときのアクセントは複合語アクセント規則が余剰的に適用されたのか，適用されなかったのかが分からない。小林ら(2015)によって PAN が規則の適用に対する示唆が得られていることを考えると，このような解釈にも可能性は十分に考えられるだろう。一方で今回の違反アクセントはどちらも「起伏型が期待されるところ

で平板型が現れた」ものであった。PAN が規則適用の逸脱により発生するならば，起伏型から（別の）起伏型への逸脱のような場合にも同様の現象が観察されるはずである。

4. 結論と課題

本発表では複合語アクセント規則の適用に関して行った ERP の計測実験について報告した。同じ複合語アクセントに対する違反であっても，適用が明白な場合には潜時 300 ミリ秒前後に ERP 成分が観察されたのに対して，適用が余剰的な場合には先行研究で報告されている ERP 成分が観察されなかった。

今回は規則適用の結果が起伏型になるものを対象としたが，日本語には複合語で平板型になる単語も一定数存在する（例：島，病，色）。これらに対する違反でも同様の成分が観察されるかなど，まだ観察の不十分な部分が存在する。また，今回は複合語のアクセント規則を対象としたが，外来語のアクセントも何らかの規則が適用されたものだと解釈するのが一般で，これについても同様の結果が期待される。

最後に実験に用いる音声について述べておく。今回の実験では操作した F0 を刺激として使用した。これは F0 の操作を行わなければ平板型と起伏型の違いがターゲットとしていた後部要素第 1 モーラから第 2 モーラにかけてではなく，前部要素の第 2 モーラですでに現れてしまうからであった。ただし，本当にこの時点で平板型と起伏型の違いを予測しているのか，その予測は言語処理において強いものなのかについては検討が行われていないので今後の課題としたい。

参照文献

Kawahara, Shigeto (2015) The phonology of Japanese accent. Kubozono, Haruo (ed.) *Handbook of Japanese Phonetics and Phonology*. Mouton.

Kubozono, Haruo (1993) *The Organization of Japanese Prosody*. (Doctoral dissertation, University of Edinburgh, 1988), Tokyo: Kurosio.

McCawley, James D. (1968) *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

Pierrehumbert, Janet B. and Mary Beckman E. (1988) *Japanese tone structure*. MA: MIT Press.

Tamaoka, Katsuo, Nobuhiro Saito, Sachiko Kiyama, Kalinka Timmer, and Rinus G. Verdonschot (2014) Is pitch accent necessary for comprehension by native Japanese speakers? An ERP investigation. *Journal of Neurolinguistics* 27: 31-40.

小林由紀，広瀬友紀，伊藤たかね (2015)「日本語複合語アクセント規則の違反に関わる ERP 研究」日本言語学会第 150 回大会.

杉藤 美代子 (1994-1999)『日本語音声の研究　第 1 巻～第 7 巻』和泉書院.

広瀬 友紀, 小林 由紀, 伊藤 たかね (2017)「言語理解におけるピッチアクセント情報：事象関連電位測定実験による検討」Prosody and Grammar Festa（国立国語研究所「対照言語学」プロジェクト 第 1 回合同研究発表会）.

安永 大地 (2017/印刷中)「言語(ERP)」, 鈴木直人・片山順一(編)『生理心理学と精神生理学　第 II 巻　応用』, pp. 91-97(予定), 京都: 北大路書房.

WS1-2 ワークショップ（1）

統語的曖昧性を解消する韻律的手段：東京方言と近畿方言[1]

五十嵐 陽介（一橋大学）・広瀬 友紀（東京大学）
y.igarashi@r.hit-u.ac.jp, hirose@boz.c.u-tokyo.ac.jp

1. はじめに

日本語東京方言において，統語論における枝分かれ構造の違いが韻律に反映されることが知られている．この統語論と韻律との関係は，「右枝分かれ構造が埋め込まれた統語境界において，基本周波数（F0）の立ち上げが起こる」という形で定式化されている[4]．この写像関係は他方言にも観察されるが[5, 9, 10]，大阪方言を中心とした近畿方言に関しては，枝分かれ構造の違いが F0 に明瞭に反映されないという報告がある[3, 7, 8]。

近畿方言において枝分かれ構造の違いが F0 に明瞭に反映されないのはなぜだろうか。その理由を近畿方言におけるアクセント体系に求める仮説（「アクセント制約仮説」）がある。この仮説によると，近畿方言は，対立するアクセント型の数が他の方言より多いので，F0 は専らアクセント型の区別に用いられ，このことが，統語構造に従って文レベルの F0 を変動させることへの制約となるという[6, 8]。ではポーズはどうであろうか。大きな統語境界には頻繁にポーズが置かれる傾向があることが知られている[2]。アクセント制約仮説はポーズに関しては何も予測しない。したがって，統語境界におけるポーズの頻度は，東京方言と近畿方言とで異ならないことが期待される。

本研究は，統語論と韻律の関係の方言差を明らかにすることを目的として，文を構成する語は同一であるが，統語的枝分かれ構造が異なる曖昧文を，東京・近畿両方言話者に発話させる産出実験を行った。

2. 手法

2.1. 話者

話者は，東京方言話者 7 名（女性 3 名）と近畿方言話者 6 名（女性 3 名）であった[2]。いずれも 19 歳から 29 歳までの大学生，大学院生であった。

2.2. テスト文

テスト文は，表 1 に示す 3 文節を，フレーム文（東京方言では「...はどこにいるの？」，近畿方言では「...はどこにおんねん？」）の冒頭に持つ 36 文であった。テスト文は，近畿方言におけるアクセント型の違いに基づいて，3 種類のデータセット（Dataset I, II, III）に分けられた。具体的には，第 1 文節と第 2 文節のアクセント型の組み合わせが，Dataset I は高起式無核＋高起式無核であり，Dataset II は高起式無核＋低起式無核であり，Dataset III は低起式無核＋高起式無核であった。第 3 文節は常に高起式無核であった。アクセント型の違いが関わるのは近畿方言のみであり，東京方言ではすべての文節が無核（平板型）であった。それぞれのデータセットは，枝分かれの異なる文のペア 6 種類から構成された。

2.3. 録音，分析手順，計測

順序をランダムにしたテスト文を被験者に提示し発話してもらった。その際，文の曖昧性解消のための説明文をテスト文とともに提示した（表 2）。発話された音声を Marantz

[1] 本稿は，国立国語研究所共同研究プロジェクト「対照言語学の観点から見た日本語の音声と文法」の助成を受けて行った成果の一部である。

[2] 「東京方言話者」は，東京，神奈川，埼玉，千葉で育ち，そこの日本語を母語として習得し継続的に使用してきた人と定義した。一方，「近畿方言話者」は，大阪，京都，奈良，兵庫，滋賀，和歌山で育ち，核と式が弁別的なアクセント体系を有する日本語を母語として習得し継続的に使用してきた人と定義した。大学進学以前に長期的な外住歴がある人や，家庭内で他言語を使用する人は除外した。

PMD660 を用いて録音した（16bit，44.1k Hz）．期待されたアクセント型とは異なる型を伴う発話を分析から除外した。東京方言 250 発話，近畿方言 156 発話が分析対象となった。

録音された音声は Praat [1]で音響的に分析した。計測点は，文節境界，文節の直後のポーズ（無音）区間，および F0 値を計測するための 3 点（文節の第 1，第 2，最終モーラのそれぞれ中央付近）であった。この 3 点から対数 F0 を抽出し，ピッチレンジの個人差を最小化するために，話者ごとに Z スコアに変換した。F0 のみあるいはポーズのみで曖昧性が解消されるか否かを検討するために，それぞれを独立変数とした判別分析を行った。

3. 結果と考察

3.1. F0 値

図 1 は，計測された点における F0 値の平均値（全話者のデータをプール）である。東京方言（左）では枝分かれの違いに起因する F0 値の差が顕著であるのに対して，近畿方言（右）では明瞭ではない。表 3-4 は，9 点の F0 値によって枝分かれの違いを判別する判別分析（強制投入法）の結果である。東京方言ではすべてのデータセットで有意な正準判別関数が得られた。また正判別率は 86.9%以上であった（Dataset I, II, III の順に 92.8%, 89.2%, 86.9%）。一方，近畿方言ではいかなるデータセットでも有意な正準判別関数は得られなかった。

3.2. ポーズ

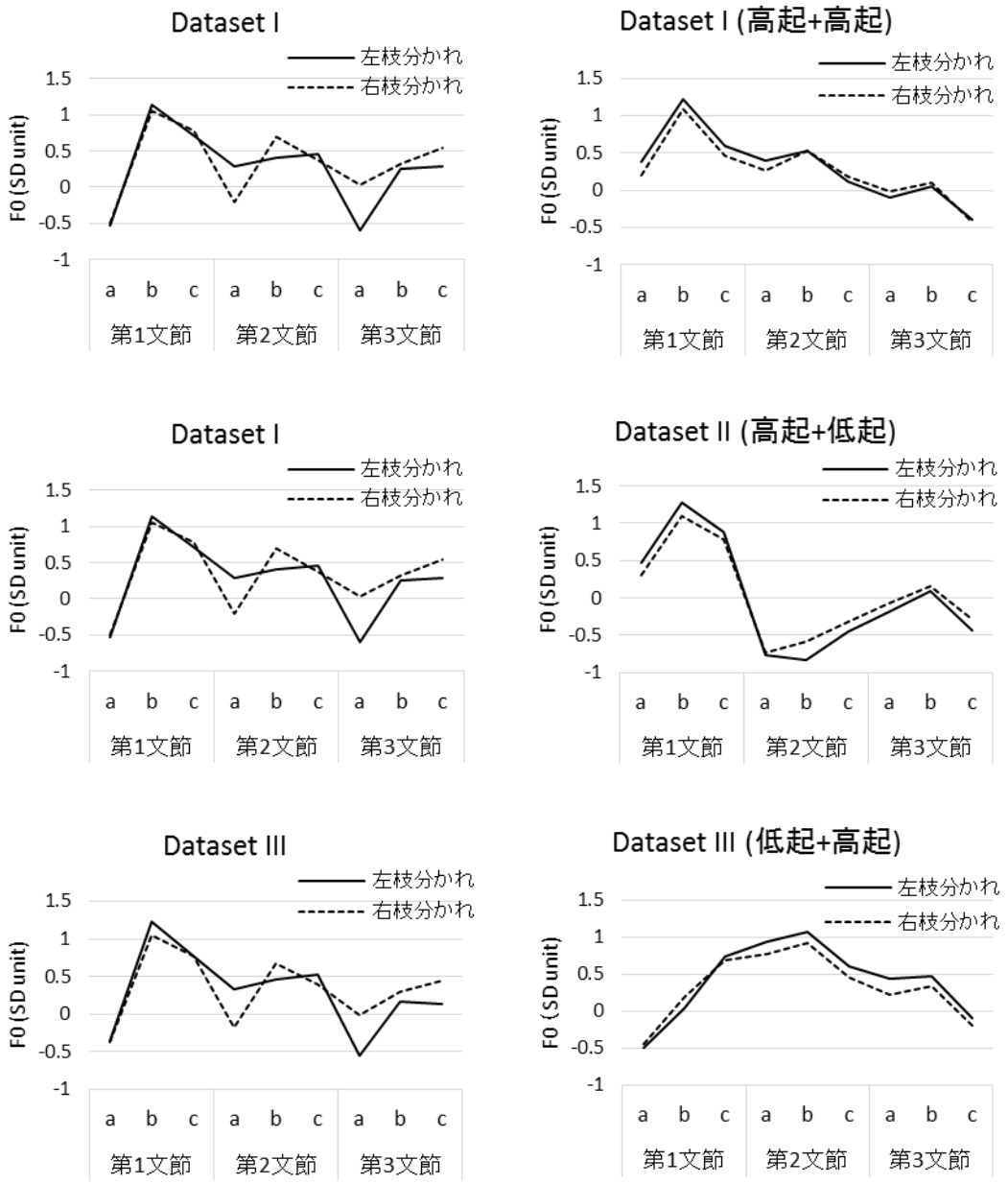
図 2 はポーズの頻度（全話者のデータをプール）を表す。両方言ともに，左枝分かれの場合は第 2 文節後に，右枝分かれの場合は第 1 文節後にポーズが置かれる傾向がある。明瞭な方言差はその頻度に認められる。全くポーズが置かれない発話が，近畿方言では全体の約 67%（全 156 発話中 105 発話）を占めるのに対して，東京方言では全体の約 20%（全 250 発話中 49 発話）を占めるに過ぎない。表 5-6 は，ポーズによって枝分かれの違いを判別する判別分析（強制投入法）の結果である。両方言ともに，すべてのデータセットにおいて有意な正準判別関数が得られた。一方，正判別率には方言差がみられる。正判別率は，東京方言では 83.3%以上であるが（Dataset I, II, III の順に 89.2%, 85.5%, 83.3%），近畿方言では 68.6%以下であり（順に 68.6%, 64.2%, 63.5%），東京方言より近畿方言のほうが低い。

表 1: テストフレーズ

	左枝分かれ	右枝分かれ
Dataset I (高起+高起)	[[草履好きの 楊枝屋の] 親父さん]	[草履好きの [楊枝屋の 親父さん]]
	[[アニメ好きの ラーメン屋の] おばさん]	[アニメ好きの [ラーメン屋の おばさん]]
	[[水菜好きの 塗り物屋の] 継母]	[水菜好きの [塗り物屋の 継母]]
	[[塗り絵好きの 煮豆屋の] 姪御さん]	[塗り絵好きの [煮豆屋の 姪御さん]]
	[[物真似好きの うどん屋の] 親方さん]	[物真似好きの [うどん屋の 親方さん]]
	[[指輪好きの ワイン屋の] 親御さん]	[指輪好きの [ワイン屋の 親御さん]]
Dataset I (高起+低起)	[[草履好きの 扇屋の] 親父さん]	[草履好きの [扇屋の 親父さん]]
	[[アニメ好きの お握り屋の] おばさん]	[アニメ好きの [お握り屋の おばさん]]
	[[水菜好きの 織物屋の] 継母]	[水菜好きの [織物屋の 継母]]
	[[塗り絵好きの 団子屋の] 姪御さん]	[塗り絵好きの [団子屋の 姪御さん]]
	[[物真似好きの おでん屋の] 親方さん]	[物真似好きの [おでん屋の 親方さん]]
	[[指輪好きの 鰻屋の] 親御さん]	[指輪好きの [鰻屋の 親御さん]]
Dataset I (低起+高起)	[[暖簾好きの 楊枝屋の] 親父さん]	[暖簾好きの [楊枝屋の 親父さん]]
	[[漫画好きの ラーメン屋の] おばさん]	[漫画好きの [ラーメン屋の おばさん]]
	[[青菜好きの 塗り物屋の] 継母]	[青菜好きの [塗り物屋の 継母]]
	[[輪投げ好きの 煮豆屋の] 姪御さん]	[輪投げ好きの [煮豆屋の 姪御さん]]
	[[ヨーヨー好きの うどん屋の] 親方さん]	[ヨーヨー好きの [うどん屋の 親方さん]]
	[[眼鏡好きの ワイン屋の] 親御さん]	[眼鏡好きの [ワイン屋の 親御さん]]

表 2: 曖昧性解消のために提示された説明文の一例.

テスト文	暖簾好きの楊枝屋の親父さんはどこにいるの（どこにおんねん）？	
説明文	左枝分かれ	楊枝屋は暖簾が好きですが，その親父さんは掛け物に興味がありません。
	右枝分かれ	楊枝屋の親父さんは暖簾が好きですが，楊枝屋自身は掛け物に興味がありません。



(a) 東京方言 (b) 近畿方言

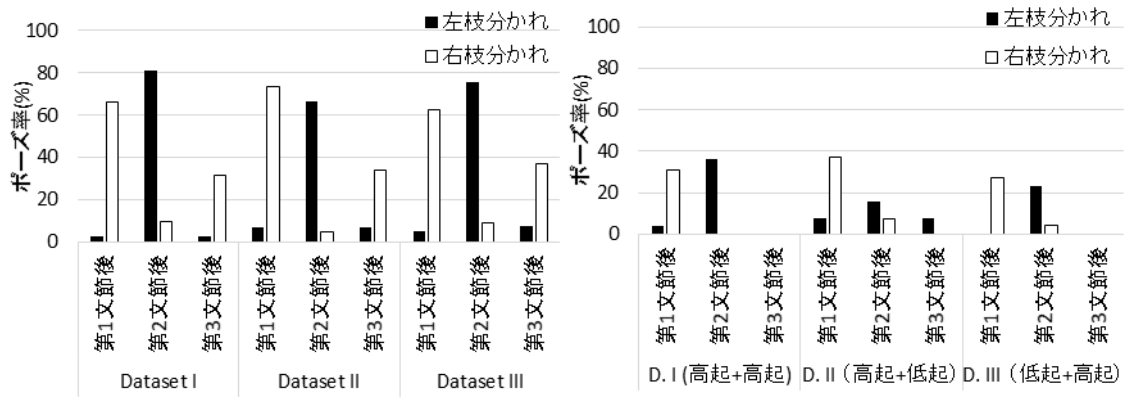
図 1: F0 値の平均値

表 3: F0 に関する判別分析の結果：東京方言

	固有値	正準相関	Wilks' λ	χ^2	df	P
Dataset I	2.142	.826	.318	87.584	9	<.001
Dataset II	1.412	.765	.415	67.360	9	<.001
Dataset III	1.137	.729	.468	58.855	9	<.001

表 4: F0 に関する判別分析の結果：近畿方言

	固有値	正準相関	Wilks' λ	χ^2	df	P
Dataset I（高起+高起）	.162	.373	.861	6.673	9	.671
Dataset II（高起+低起）	.260	.454	.794	10.731	9	.295
Dataset III（低起+高起）	.187	.397	.842	7.801	9	.554



(a) 東京方言　　(b) 近畿方言

図 2: ポーズ率

表 5: ポーズに関する判別分析の結果:東京方言

	固有値	正準相関	Wilks' λ	χ^2	df	P
Dataset I	2.219	.830	.311	92.952	3	<.001
Dataset II	1.405	.764	.416	69.779	3	<.001
Dataset III	1.182	.736	.458	62.819	3	<.001

表 6: ポーズに関する判別分析の結果:近畿方言

	固有値	正準相関	Wilks' λ	χ^2	df	P
Dataset I（高起+高起）	.405	.537	.712	16.325	2	<.001
Dataset II（高起+低起）	.197	.406	.835	8.916	3	<.001
Dataset III（低起+高起）	.258	.453	.795	11.261	2	<.001

4. 結論

統語的枝分かれ構造が異なる曖昧文を，東京・近畿両方言話者に発話させる産出実験を行った。第 1 に，先行研究の実験結果を再現し，近畿方言では統語的枝分かれ構造が F0 によって明瞭に区別されないことを示した。第 2 に，近畿方言では，枝分かれ構造の違いがポーズによっても必ずしも区別されないことを示した。近畿方言において左右枝分かれ文の間の F0 の差が小さい事実は，アクセント制約仮説によって説明することができる。しかしながらポーズの頻度が近畿方言において低いことを示した結果は，アクセント型の数とは無関係に，統語論と韻律の写像規則に方言差が存在していることを示唆している。

参考文献

[1] Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.28, retrieved from http://www.praat.org/.／[2] Cruttenden, Alan (1997). Intonation. Second edition. Cambridge: Cambridge University Press.／[3] Igarashi, Y. (2014). 'Typology of intonational phrasing in Japanese dialects', In: Sun-Ah Jun (ed.) *Prosodic Typology II*, pp. 464-492. New York: Oxford University Press.／[4] Kubozono, H. (1988). The Organization of Japanese Prosody, Ph.D. dissertation, Edinburgh University.／[5] 五十嵐陽介(2010). 「統語論における枝分かれ構造は韻律にどのように反映されるのか?: 近畿方言と東京方言の場合」『第 24 回日本音声学会全国大会予稿集』: 185-190.／[6] 上村幸雄(1997).「日本語音声の歴史的なふかさと地域的なひろがり」杉藤美代子(監), 佐藤亮一・真田信治・加藤正信・板橋秀一(編)『日本語音声 1 ：諸方言のアクセントと イントネーション 』, pp. 21-61, 東京 : 三省堂.／[7] 郡史郎 (2006).「韻律特徴の地域差」, 広瀬 啓吉(編著)『韻律と音声言語情報処理―アクセント・イントネーション・リズムの科学』, pp. 50-64, 東京 : 丸善. 郡史郎 (1989). 「発話の音調を規定する要因－日本語イントネーション論」『吉沢典男教授追悼論文集』(東京 : 東京外国語大学) , pp. 116-127.／[8]杉藤美代子(2001). 「文法と日本語のアクセントおよびイントネーション－東京と大阪の場合」音声文法研究会（編）『文法と音声』, pp. 197-210. 東京 : くろしお出版.／[9]前川喜久雄(1997). 「アクセントとイントネーション―アクセントのない地域」杉藤美代子(監), 佐藤亮一・真田信治・加藤正信・板橋秀一(編)『日本語音声 1 ：諸方言のアクセントと イントネーション 』, pp. 97-122. 東京 : 三省堂.／[10]松浦年男(2016).「二型アクセント方言のイントネーション」日本音韻論学会（編）『現代音韻論の動向 : 日本音韻論学会の歩みと展望』, pp. 100-103. 東京: 開拓社.

日英のコーパスを用いたプロソディ研究*

北原 真冬（上智大学外国語学部）
mafuyu@sophia.ac.jp

1. はじめに

本発表の射程は(1)英語母語話者，(2)日本語母語話者，そして(3)日本語を第一言語とする英語学習者，のプロソディ研究である．実は(3)を対象にしようとすると，必然的に(1)をターゲット，(2)をベースラインとして押さえておかねばならない．研究プロジェクトの規模を無闇に拡大しないために，特に(1), (2)についてコーパスを用いた下調べは必須である．単語頻度・親密度データベースとして(1)に Hoosier Mental Lexicon (HML), (2)に日本語の語彙特性(Psylex)，自然発話コーパスとして(1)に Buckeye Corpus，(2)に Corpus of Spoken Japanese (CSJ)を用いた研究のデザインを粗描する．

2. 単語頻度・親密度データベースの利用

音声の産出や知覚について実在する単語を用いて実験をする場合，実験の焦点となる変数以外についてはなるべく平等な条件を揃えたい．例えば，単語アクセントの有無と句頭のピッチの関わりを焦点とする実験を組むには，無アクセントの単語群と有アクセントの単語群を用意することになる．この時，二つの単語群の親密度は一定の基準範囲に収まるように揃えることが望ましい．どちらかの単語群になじみのない単語が含まれていると，結果にバイアスが入ってしまう恐れがある．また，各単語の近傍(neighborhood)の密度や頻度のバランスも考慮することが望ましい．以下では，それらの統制に必要な英語と日本語の単語親密度データベースを概観する．

2.1. Hoosier Mental Lexicon (HML)

HML (Nusbaum, et al. 1984)はインディアナ大学心理学部において開発された単語親密度データベースである．Webster Pocket Dictionary (1964)の見出し語約 2 万項目について，参加者が 7 段階の尺度で親密度を評定したデータが, 単語の音韻表記等と共に収録されている．

表 1. HML 所収データのサンプル

正書法	音韻表記	頻度	親密度	内容/機能語	品詞
a	x	23237	7.00000	f	=nr
aback	x-b'@k	2	5.08333	c	a
abacus	'@<bx-kxs	0	5.16667	c	n

* 本発表は国立国語研究所共同研究プロジェクト「対照言語学の観点から見た日本語の音声と文法」の一部である．また，田嶋圭一氏(法政大学)，米山聖子氏(大東文化大学)との共同研究の成果を含む．関係各位に感謝する．

2.2. Psylex

日本語の語彙特性(天野・近藤, 1999)は NTT コミュニケーション科学基礎研究所において開発された単語親密度を中心としたデータベースである．新明解国語辞典第四版の見出し語約 7 万項目について，HML と同様の 7 段階尺度を用いて，音声だけでなく表記についての親密度と音韻表記等を収録している．他にも，朝日新聞約 14 年分のデータにおける単語の出現頻度，単語の心像性についての評定値，3 万語の増補も併せた総計 9 巻に及ぶデータが三省堂から出版された. ID 番号によって 1,3,7 巻から情報を集約したサンプルは以下の通りである．

表 2. Psylex 所収データのサンプル

ID	カナ	表記	長さ	アクセント	ローマ字	頻度	AV 親密度	A 親密度	V 親密度
434930	ハシ	端	2	0	ha.shi	2440	5.125	5.906	5.156
434940	ハシ	嘴	2	1	ha.shi	21	2.219	5.500	2.938
434950	ハシ	箸	2	1	ha.shi	277	5.625	5.500	5.250
434960	ハシ	橋	2	2	ha.shi	3975	6.250	5.906	6.062

3. 自然発話コーパスの利用

3.1. Buckeye コーパス

Buckeye コーパスは米国オハイオ州立大学の Mark Pitt らをリーダーとするチームによって作成された自然発話コーパスである．オハイオ州コロンバス在住の成人男女 40 名を対象としたインタビュー形式で，およそ 30 万語分の発話が収録されている。音声データは，まず自動認識処理によってラベル付けが与えられたのち，訓練を受けた音声学者が修正した．

Buckeye コーパスは各話者の録音が数個のセッションに分けられており, sXXYYZ という形式のフォルダに各セッションが収められている．ここで XX は話者番号，YY はセッション番号，Z はセッション記号であり，例えば s0101a は話者番号 01，セッション 01a という意味である. 各フォルダは音声ファイル(sXXYYZ.wav), 書き起こしテキスト(sXXYYZ.txt), 及び音声解析ソフトウェア Praat のアノテーション(sXXYYZ.TextGrid)を含んでいる．

Praat のサウンドエディターで wav ファイルと TextGrid ファイルを開くと図 1 のような画面表示となる．1 番上のパネルは音声波形，2 番目はスペクトログラム，その下のテキストの入った枠は，単語層，分節音層，コメント層が順に並んでいる．

単語層は全て小文字を用いているが，基本的に英単語のスペルそのままであり，語末にセミコロンが付されている．また図 1 の冒頭部にあるように「<HES>itation=ためらい」など，いくつかのタグが用いられている．分節音層は DARPA phonetic alphabet に準拠した音声表記であり，IPA などの特殊フォントを用いていない．使用されているタグ，記号，コメントなどの一覧は付属するマニュアルに記載されている．

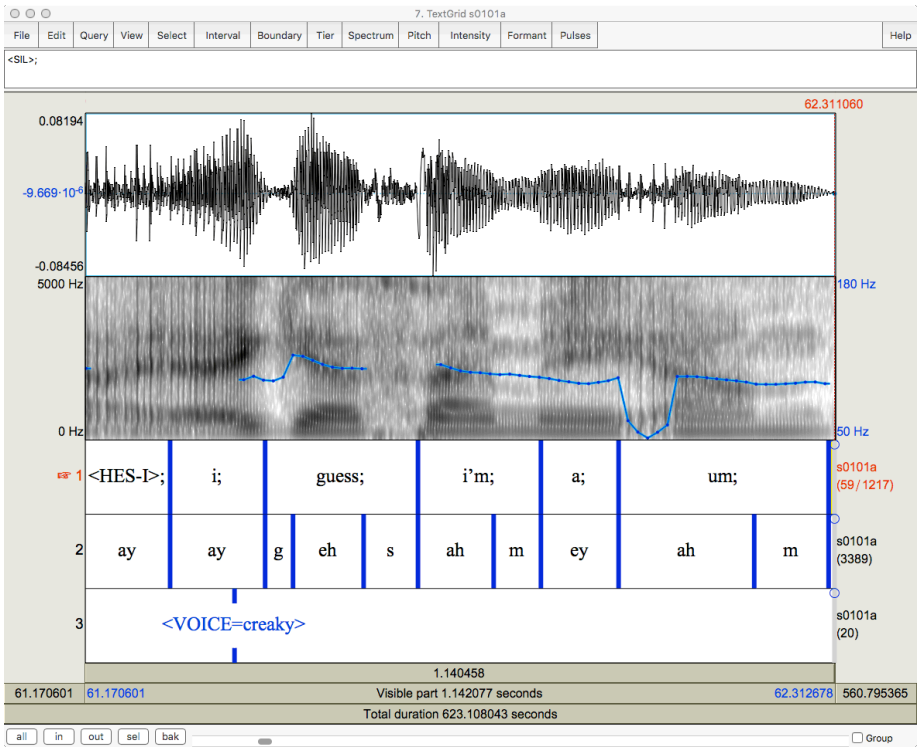


図 1: Buckeye コーパスを Praat で開いたところ

ここでは，Buckeye コーパスにおける歯茎弾音(alveolar flap)の生起についての調査を分析の具体例として示す．歯茎弾音は北米の英語において，/t/, /d/が強勢のある母音と次の母音に挟まれた時に生起する．例えば，*better, rider* のように語中で起こるだけでなく，*get up, need it* のように単語間でも起こる．Praat の TextGrid でデータが提供されているため，比較的簡単な Praat スクリプトによって例えば以下のようなデータを得ることができる．

表 3. Buckeye コーパスにおける検索のサンプル

単語	検索対象	持続時間(秒)
thirty;	dx	0.020144999999999413
created;	dx	0.026443999999997914

筆者を含む研究グループでは，これに基づき，帰国子女を含む様々なグループの日本人英語学習者の弾音化と比較することで，英語学習における非音素的な特徴の役割を解明しようとしている．その際，コーパスの探索は実験のベースラインを与えてくれるという意味で極めて重要である．いわゆる実験室発音(lab speech)ではない，自発的な会話の音声データを検索し，任意の音声現象について環境ごとの出現率が得ることで，より特定的な条件や狙いを定めた単語群についての発話産出実験や知覚実験をデザインできる．

3.2. 日本語話しことばコーパス

日本語話し言葉コーパス(Corpus of Spoken Japanese: CSJ)は，国立国語研究所，情報通信機構，東京工業大学の 3 者が共同で開発した，大規模な自発音声のデータと詳細なアノテーションを持つコーパスである．Buckeye コーパスに比べると，形態・統語情報，音声・韻律情報などが豊富に用意されている．ここでは北原・米山(2014)において扱った，母音の持続時間に関する後続子音の有声性の影響を一例として紹介する．

英語の母音が有声子音の前では無声子音の前よりも約 50%も持続時間が長いことはよく知られている．しかし，有声子音の前で母音が長くなるのは，調音的に見て自然な同化プロセスであり，日本語の発話においても英語ほどではないにしても，似たような傾向が見出せるかどうかは調査に値する．まず日本人乳幼児の発話データベース(Amano et al., 2008)を調べると，後続子音の有声性による母音長への影響が見られた．つまり，調音的に自然なプロセスが乳幼児の日本語発話においては現れる．ところが成人になると，後続子音の影響は弱まり，短母音の長さにほとんど差がない場合も多いことが CSJ の分析から分かった．ここでの CSJ の利用法は，Buckeye コーパスの場合と基本的には変わらず，コーパス内にあるアノテーションのみを用いている. Praat の TextGrid に対するスクリプトを用いれば，任意の分節音についての持続時間やその周辺のラベルを抽出することは容易である．

4. まとめと展望

本発表では主に 4 つのデータベース・コーパスを扱い，それらを元に日本語および英語のプロソディ分析に役立つ情報の取り出し方について触れた．これらは分析の基礎的作業だが，スクリプト言語によるプログラミングを含むため，一般には敷居が高く感じられることも多い．今後，なんとか敷居を下げる方策も探していきたいと考えている．

参考文献

天野成昭・近藤公久 (1999)『NTT データベースシリーズ:日本語の語彙特性』東京：三省堂

Amano, S., K. Tadahisa, and K. Kato (2008) "Development of NTT Infant Speech Database," *Technical report of IEICE*, 108(59), 29-34.

Boersma, P. and D. Weenink (2017) *Praat: doing phonetics by computer* [Computer program]. Version 6.0.29, retrieved 24 May 2017 from http://www.praat.org/

北原真冬・米山聖子 (2014) 「後続子音による母音長の変化：幼児・成人の日本語コーパス分析と成人の英語学習データ」*JELS* 31, 44-48.

国立国語研究所 (2006) 『日本語話し言葉コーパス』[pj.ninjal.ac.jp/corpus_center/csj/]

Nusbaum, H. C., D. B. Pisoni, and C. K. Davis (1984) "Sizing up the Hoosier Mental Lexicon," *Research on Speech Perception Progress Report* 10, 357-375. Bloomington: Indiana University

Pitt, M. A., L. Dilley, K. Johnson, S. Kiesling, W. Raymond, E. Hume, and E. Fosler-Lussier (2007) "Buckeye Corpus of Conversational Speech (2nd release)," [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor)

ワークショップ（2） 趣旨説明

英語教員を目指す人への英語音声指導法

山根 典子 （広島大学）
yamanen@hiroshima-u.ac.jp

2017 年 5 月、日本音声学会では学習指導要領に定める英語音声教育実現のため「音声に関する科目の履修を英語の教育職員免許状取得の必須条件とすること」を提言している。現在の学校教育の現場では、中学校、高等学校の英語科目はもとより、小学校の外国語活動も 2018 年度から教科化されることになり、英語の音声に関して、今後ますます教師の理解と実践が強化されなければならないという状況にある。

さて音声学が教職の必須条件となった場合、履修した学生は、IPA や音声・音韻に関する専門用語と共に、体系的な知識を身につけていくが、知識・発音力・指導力が全く別ものであることにすぐに気づくかもしれない。また晴れて音声学を履修し終え、教員免許を取得し、教壇に立った時には、何も知らない学生を目の前に、どう英語嫌いを増やさずに発音力を身につけさせようかと方法論を模索することも考えられる。

本ワークショップでは、まずはなるべく専門用語を使わずに、比較的手軽に教室で使える英語音声の指導法を、母音、子音、プロソディーに分類して紹介する。

野北講師は、一般米語全 13 母音の指導法を紹介する。5 母音の区別しか馴染みがない日本人にとって 13 母音を覚えるのは大変なことだと思うかもしれないが、全体像の把握を容易にするため、日本語の音を利用した一覧表を作成した。それゆえ 日本語に区別の無い/ɛ/と/æ/や /ɑ/と/ʌ/ の違いも、日本語母語話者に馴染みのある形での習得が実現する。学習者が覚えやすいように 13 母音を歌にのせることも提案されている。

桑本講師は、英語教員に求められる子音の理解と実践方法についてを紹介する。桑本講師は、教員免許更新講習、および高等専門学校や大学において発音指導の授業実践の実践から、英語発音の入門書を出版された。今回はその中から、日本語に区別の無い 1) [l] と [ɹ] の区別 2) [wʊ] [jɪ] のような「わたり音 + 高母音」の連続 3) 歯擦音 [θ] [s] [ʃ] の区別を取り上げ、効果的な発音教授法を示す。発音に役立つアプリの開発も手がけている。

ティーマン講師と山根は、プロソディーを習得させるための触覚指導法「ハプティック」を紹介する。ティーマン講師は日本での 20 年の英語指導の経験から、学習者の発音指導に身体行動を取り入れている。イントネーションに Touchinami, フォーカス語や強勢リズムに Rhythmic Fight Cub, 強勢音節・無強勢音節に Syllable Butterfly という触覚指導法を示す。パイロットスタディーの結果を、体験談、音声分析や聴覚判断で示す。

いうまでもなくこれらは音声指導法の一部にすぎないが、聴衆の方たちにも指導法を体験していただき、教師の役割、理論と実践のバランス、英語音声指導法研究の方向性などについて、意見交換ができれば本望である。

一般米語全 13 母音の指導法: How to teach all 13 vowels of General American English to Japanese people

野北 明嗣（国士舘大学）
anogita@kokushikan.ac.jp

1. はじめに

一般に、日本語には母音が 5 個しかなく、米英語（以下、英語）にはその倍以上あるから、英語の母音は難しいという「通説」がある。しかしそれは大きな誤解である。なぜなら、音韻学者にとっては常識であるように、音素の数は、数え方次第でいくらでも変わるからである。そこで、本稿では以下の 2 点にフォーカスを当てる。

- 英語の母音は、決して日本語のア、イ、ウ、エ、オのような短い単母音が 13 個あるわけではなく、長さと二重母音性が区別のカギになる。
- 日本語の五十音表のように整頓された表で、英語の母音の全体像を把握する。

2. 英語の母音の音声的特徴に関する盲点

英語母語話者は、英語の母音を「長さ」と「二重母音性」によって聞き分ける。この英語音声学では常識のような事実が、日本の英語教育でほとんど応用されていないのは残念である。まず英語の母音には intrinsic duration があり、Hillenbrand et al. (2000)によれば、各母音の平均持続時間は、図 1 に示されているとおり長いものと短いものがある。

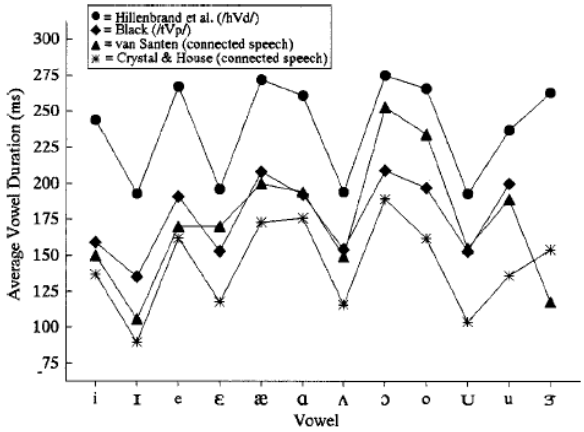


図 1: 米英語の母音の平均持続時間 (Hillenbrand at al., 2000, 3014)

図を見ればわかるとおり、例えば[ɑ]（*hot* の母音）は長い母音で、[ʌ]（*hut* の母音）は短い母音である。Hillenbrand at al. (2000)によれば、[ʌ]を人工的に長くしたものを英語母語話者に聞かせると、[ɑ]と聞き間違えられる傾向があり、逆に[ɑ]を短くしたものは[ʌ]と知覚される傾向があるという。つまり、長さが母音の区別に重要な役割を果たしているのである。

さらに、日本語のア、イ、ウ、エ、オが文字通り単母音であるのと違って、英語でいわゆる「単母音」と言われているものの多くは、実際には二重母音がかっている。多くの北米の方言で、例えば *bid*, *bed*, *bad*, *bud*, *good* はそれぞれ [bɪəd], [bɛəd], [bæəd], [bʌəd], [gʊəd] の

ように、二重母音的に発音される (Andersen, 1972)。

これらを踏まえた上で、Nogita and Lin (2016)によれば、日本語の 1 拍及び 2 拍の単母音や母音連続をカナダ英語母語話者に聞いてもらい、英語のどの母音に聞こえるかという知覚実験を行ったところ、日本語の母音を組み合わせていけば、英語の全 13 母音に相当する音を作ることができることがわかった（後述の表 1 参照)。つまり日本人にとって英語の全ての母音を発音し分けることは、決して難しいことではないのである。（ただし native-like な発音は難しい。）子音の L や TH 等と違い、母語に既に対応する音韻カテゴリー（音声的には全く同じとは限らない）が存在することは、学習者の負担を大きく軽減する。

3. 米英語の母音一覧表を把握する

日本語の諸方言同様、英語も母音音素の数には方言差があるが、多くの北米英語の方言では 13 個であり(Labov, Ash, & Boberg, 2006)、ここでもそれを採用する。ここで、各言語の音の「目録」という概念を理解する必要がある。例えば日本語では、母語話者、非母語話者を問わず、学習の初期段階で五十音表を学び、文字と同時に音の体系を覚える。そして大雑把に言えば、日本語は五十音表の音しか使わないので、五十音表の音さえ覚えてしまえば、理論上は全ての日本語の単語を発音できるわけである。英語も同様で、母音に関して言えば、表 1 に示してある 13 個の母音しか使わない。

ところで、日本語の漢字の読みを表記するのには、ひらがな、カタカナ、ローマ字、発音記号等、いくつかの表記方法がある。同様に、英語の発音を表記する方法も複数ある。まず英和辞典や教科書でお馴染みの発音記号。他には、日本人にはあまり馴染みがないかもしれないが、英語圏で使われる phonetic spelling（読み通りのつづり）がある。これにも大きく分けて 2 種類あり、例えば *Germany* を **jur**-muh-nee (c.f., Scholastic Pocket Dictionary) のように補助記号を使わず表記するものと、例えば *plane* と *plan* をそれぞれ plān、plăn (c.f., The American Heritage Dictionary)のように補助記号を使って表記するものがある。この機会に多くの日本人に phonetic spelling を知ってほしい。表 1 では、極力 1 音素 1 文字に近づけるため補助記号を使った phonetic spelling を採用し、それに対応するジーニアス英和辞典の発音記号、さらに前述の Nogita and Lin (2016)の知覚実験による日本語の対応する音を加えた。

表 1: 北米英語 13 母音一覧表

	Long/Alphabet vowels				
Phonetic spelling [a]	ā	ē	ī	ō	o͞o
対応する発音記号 [b]	[eɪ]	[i:, i]	[aɪ]	[oʊ]	[u:, u]
対応する日本語音 [c]	エイ	イー	アイ	オウ	ウー
単語例	*mate*	*mete*	*site*	*note*	*root*
	Short/Relative vowels				
Phonetic spelling	ă	ĕ	ĭ	ŏ	ŭ
対応する発音記号	[æ]	[ɛ]	[ɪ]	[ɑ:]	[ʌ]

対応する日本語音	エア	エー	エ	アー	ア
単語例	*mat*	*met*	*sit*	*not*	*cut*
	Vowel digraphs				
Phonetic spelling	oi	ou	o͝o		
対応する発音記号	[ɔɪ]	[aʊ]	[ʊ]		
対応する日本語音	オイ	アオ	ウ		
単語例	*coin*	*loud*	*book*		

[a] The American Heritage Dictionary の発音表記。 [b] ジーニアス英和辞典の発音記号。 [c] Nogita and Lin (2016)の知覚実験による日本語の対応する母音。

これらの母音を全て覚え、子音一覧表も同様に覚えれば、理論上は全ての英単語を発音できるわけである。ただし、日本語の五十音表を覚えるのにはそれなりの時間と労力を要するように、英語の音の一覧表を覚えるのにも同等の時間と労力が必要である。

4. 3 種類の母音 (Long/Alphabet Vowels, Short/Relative Vowels, Vowel Digraphs)

英語のアルファベットは A から Z まで 26 文字あるが、母音字は A, E, I, O, U の 5 文字である。そして 5 文字それぞれに 2 通りの読み方、計 10 通りの読み方がある。これはちょうど日本語の漢字に、音読みと訓読みの 2 通りの読み方があるのに似ている。つまり、13 母音のうちの 10 個は A, E, I, O, U の文字で書き表せる。残りの 3 つは、例えば子音字の SH や CH 等 2 文字で 1 音を表す「二重字(digraph)」があるが、これの母音版、OI, OU, OO で表す。

4.1. Long 読み/Alphabet 読み

A, E, I, O, U の 2 通りの読み方のうちの 1 つ目は、Long 読み (c.f., Labov et al., 2006) とか Alphabet 読み (c.f., Gilbert, 2012) と呼ばれる。Long/Alphabet 読みは、A, E, I, O, U をアルファベット名で、つまり順にエイ、イー、アイ、オウ、ユーと読む読み方である（例、*mate*, *mete*, *site*, *note*, *cute*）。Long/Alphabet 読みの A, E, I, O, U は、ā, ē, ī, ō, ū とマクロンを付けて表記されることもあり、本稿でもこれを採用する。ただし、ū はユー(/ju:/)と読み、Y の子音が入っている。そこで、本稿では ū の表記は避け、The American Heritage Dictionary に倣って、母音ウー(/u:/)だけを表す o͞o（*food* や *zoo* の母音）の表記を採用する。

4.2. Short 読み/Relative 読み

A, E, I, O, U のもう一つの読み方は、Short 読み (c.f., Labov et al., 2006)とか Relative 読み (c.f., Gilbert, 2012)と呼ばれる。日本では「フォニックス読み」と呼ぶ人もいる。Short/Relative 読みは、アルファベット名とは全く別な読み方をし、表 1 に書いた通り、それぞれ概ね日本語音のエア、エー、エ、アー、アに対応する（例、*mat*, *met*, *sit*, *not*, *cut*）。Short/Relative 読みの A, E, I, O, U は、それぞれ ă, ĕ, ĭ, ŏ, ŭ とブリーブをつけて表記されることもある。本稿でもこれを採用する。

ところで、ă, ŏ, ŭ の 3 つは、日本の英語教育では、しばしば全て日本語のアに対応すると

教えられ、ア系母音などと呼ばれたりするが、これはとんでもない誤解である。前述の Nogita and Lin (2016)の知覚実験で表されるとおり、ă, ŏ, ŭ はそれぞれ日本語音のエア、アー、アという 3 つの音韻カテゴリーに対応する。図 2 にこれをまとめる。

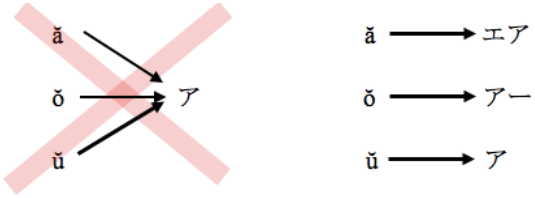


図 2: いわゆるア系母音の実際の対応

4.3. Vowel Digraphs（二重字）

残りの 3 つは、しばしば oi, ou, o͡o（又は oy, ow, o͡o）で表され、順に *coin*, *out*, *good* の母音である (c.f., The American Heritage Dictionary)。o͡o に関しては、o͞o と区別するため、本稿では The American Heritage Dictionary に倣い、前者にブリーブ、後者にマクロンをつける。

Nogita and Lin (2016)の知覚実験では、o͡o は日本語音ウに対応したが、実際は o͡o と o͞o は音質がかなり違うため、経験的には、「口を丸める o͞o に対し、o͡o はあまり口を丸めないウ」又は「o͡o は口を丸くしてア」という説明が効果的だが、今後科学的な実証が必要である。

5. 具体例

最後に、実践でこれらの母音が英単語でどのように使われているかを見ていく。英単語が日本語の語彙に入って、カタカナ発音になることは有名だが、逆に日本語の単語が英語の入るとどうなるかは、あまり注目を浴びていない。そこでここでは、日本語語源の英単語の読み方を見てみる。ハイフンは音節の分かれ目、太字は強勢を示す。

kē – **yō** – tō(dō)　京都	**tō** – kē – ō / tōk – yō　東京	**hŏn** – dŭ　本田
toy – **yō** – tŭ(dŭ)　トヨタ	**nē** – sŏn　日産	tō – **shē** – bŭ　東芝
sŏ – kē　日本酒	kŭ – **rŏ** – tē(dē)　空手	**mē** – sō - so͞op　味噌汁
kĕ – rē – **ō** – kē　カラオケ	**so͞o** – shē　お寿司	so͞o – **nŏ** – mē　津波

参考文献

Andersen, H. (1972). Diphthongization. *Language*, 11-50.

Gilbert, J. B. (2012). *Clear Speech Teacher's Resource and Assessment Book: Pronunciation and Listening Comprehension in North American English*. Cambridge University Press.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013-3022.

Labov, W., Ash, S. & Boberg, C. (2006). *The Atlas of North American English*. Berlin: Mouton de Gruyter.

Nogita, A. & Lin, H. (2016, May). *Establishing Counterparts of Japanese and English Vowels.* presented at SFU Pronunciation Mini-Conference at Simon Fraser University, BC.

WS2-2 ワークショップ（2）

英語教員に求められる子音の理解と実践方法[1]

桑本　裕二（公立鳥取環境大学）
kuwamoto@kankyo-u.ac.jp

1. はじめに

本発表は，桑本 (2017) の内容に基づき，特に子音の指導法を考察するものである．桑本 (2017) では英語の 24 子音を，初学者が理解しやすい順に 3 種に分類し，順を追って学べるようにした．本発表はそのなかから特に興味深い，また重要性が高いと思われる指導法について 3 項目選び，考察する．

2. 桑本 (2017) による英語子音の分類

桑本 (2017) では，英語の 24 子音を以下の 3 段階に分類した．

(1)　1) 日本語の音とほとんど同じ音（6 子音）

[p] [b] [k] [g] [ʧ] [ʤ]

2) 日本語と少し違う場合のある音（9 子音）

[m] [n] [j] [w] [s] [z] [t] [d] [h]

3) 日本語にない音（9 子音）

[f] [v] [θ] [ð] [ʃ] [ʒ] [l] [ɹ] [ŋ]　　桑本 (2017: 41f.)

子音の発音実践は 1) の 6 子音から行うが，「[p] は無声両唇閉鎖音である」といったような調音音声学的な言及はなるべく避ける．これらの 6 子音は，いわゆる英語特有の（そして日本語には存在しない）音ではないということについて，日本語の五十音表の「〜行」の頭子音と同じであることを認識させる[2]．

(2)　[p] ……「パ行」の頭子音　　[b] ……「バ行」の頭子音
[k] ……「カ行」の頭子音　　[g] ……「ガ行」の頭子音
[ʧ] ……「チャ行」の頭子音　　[ʤ] ……「ジャ行」の頭子音

(1) の分類中，2) と 3) にある子音については，それぞれ教授法に工夫が必要である．次節では桑本 (2017) に掲載された例に基づき，３つの項目を取り上げ，効果的な発音の教授法について考察する．

3. 実践例

3.1 [l] と [ɹ] の区別

初学者向けには，英語特有の「むずかしい」発音として [l] と [ɹ] の区別が代表的である

[1]本発表の内容は，平成 29 年度教員免許状更新講習（2017 年 8 月 4 日，公立鳥取環境大学（桑本・中村 2017)）および英語授業（「英語 LL 演習」秋田工業高等専門学校（桑本 2012)，「Intensive English」公立鳥取環境大学）における実践をふまえたものである．

[2] いわゆる「カタカナ読み」は英語発音の弊害であると言われるが，1) の子音に限っては「カタカナ読み」が正しいということを強調する．そのことによって，たとえば 3) の子音との区別を明確に意識させることができる．

ので慎重に扱うべきである点を強調する．また，[ɹ] は原則的に文字 <r> で書くが，日本語のラ行音をローマ字で書くとき，同じく <r> を用いることに影響されて，同一音であるという誤解に陥らないように注意させるべきである．また，日本語のラ行音の子音 [ɾ] が，英語の [l] とも [ɹ] とも異なる音であること，英語をはじめ，多くの言語で，流音に 1 系の音と r 系の音の区別があるのに，日本語は r 系のみしかない点が，英語発声時における [l] と [ɹ] の区別の意識を薄れさせていることなどを強調して指導するべきである．

[l] については，イラスト（図 1），調音器官の断面図（図 2）などにより舌の位置を確認した後，*lily, love* など，[l] を（主に語頭に）もつ単語を発音する練習をさせる．また，dark l [ɬ] に関しても，その舌の位置の違いや，[ɬ] が母音に近い音色をもつことが，[l] に比べて口腔が広くなるのでソノリティが高くなることに起因することを図解とともに示し（図 3），語末に [ɬ] をもつ *beautiful, full* などの単語を発音して確認させる．

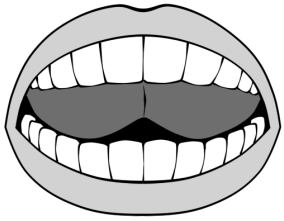


図１

図２

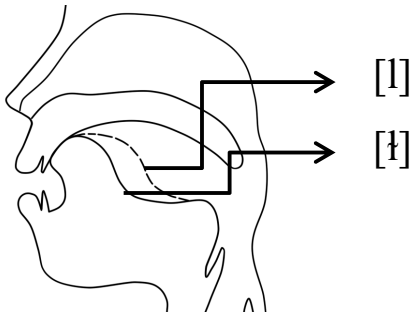


図３

[ɹ] については，以下のような記述によって説明を試みている．

(3) 「ラリルレロ」で 5 回舌が上あごをたたくのを、無理矢理一度も触れないようにして「ラリルレロ」と言おうとしてみてください。何とも言えない、柔らかいものが漂うような不思議な音色になると思いますが、このときにほぼ英語の [rarirurero][3] に近い音が出せているはずです（桑本 2017:74f.）。

(4) 日本語で「ラリルレロ」と言うと、舌の先が口の中の上を 5 回軽くたたきます。この「たたき」をなくして「ゥアゥイゥウゥエゥオ」のようになめらかな音にすると、それが英語の [r][4] です。“Laura” さんは、[l] も [r] もでてくるので良い練習になります（中村他 2015:49）。

さらに，[l] と [ɹ] の最小対立の組 (5)，[l] と [ɹ] を両方含む語 (6) をいくつか発音する演習を行う．

(5) light — right
glass — grass
low — row
lock — rock　桑本 (2017:75)

(6) rule
really
Australia
vocabulary　桑本 (2017:76)

[3] 桑本 (2017), 中村他 (2015) では，英語の [ɹ] は [r] と表記することに統一している．
[4] 注 3 を参照．

3.2 わたり音 + 高母音 の連鎖

前舌の高母音 [i] とわたり音 [j]，後舌の高母音 [u] とわたり音 [w] の類似性を，(7) のような現象から納得させ，それゆえに [ji] [wu] の連鎖が困難であることを説明する．[l]/[ɹ] の区別に比べると目立たないが，実際には潜在的に困難な英語発音であることを強調する．

(7) 「い，あ，い，あ，……」i.a.i.a.i.a… → 「や，や，……」ja.ja.ja…
「う，あ，う，あ，……」u.a.u.a.u.a… → 「わ，わ，……」wa.wa.wa…

この項目については，発表者は以下のように説明している．

(8) [wu] としては、woman [wómən]「女性」wood [wód] 「木材」[ji] としては year [jíə*r*] 「年」がありますが、いずれも外来語的にカタカナで書かれると、「ウーマン」「ウッド」「イヤー」のように、母音で始まっているように発音され、そして書かれます。……[wu] は、……「ウ（ー）」のようになってしまいます。これを防ぐために、「ウオ」をカタカナ読みするようにすると、「え？　それでいいの？」という感じですが、十分本当の音に近いといえます。“woman [wómən]” は「ウーマン」よりは「ウオマン」と読むつもりくらいの方が、ネイティブスピーカーには通じるのです。[ji] に対しては、……「イ」を強めに発音する、という感じ、たとえば、小さい「ィ」をはじめに読むつもりで、 “year [jíə*r*]” などは「ィイアー」のように読むつもりで発音すればどうでしょうか？（桑本 2017:57f.）

3.3 歯擦音 [θ] [s] [ʃ] の区別

[s] は，第 2 節の (1) の分類では，2) 日本語と少し違う場合のある音 である．1) 日本語の音とほとんど同じ音の子音群になぞらえれば，「「サ行」の頭子音」を表すことになるが，日本語では [i] が後続する場合，口蓋化して [ɕi]（し）になるので，口蓋化しない [si] の発音は，特に英語初学者にとっては極めて困難である．例えば現代日本語において，「シリーズ（series)」に対して「スィリーズ」とでも表記すべき非口蓋化音が未だに許容されていないことは，当該の [si] の発音をさらに困難にさせている一因であると推察される．

なお，英語の後部歯茎音 [ʃ] は原則的に <sh> と綴り，日本語の「し」はローマ字（ヘボン式）で <shi> と綴ることから両者の音を同一視する傾向も，（少なくとも初学者には）見受けられる（桑本・中村 2017）．桑本 (2017:85) では，英語における [sɪ/si:], [ʃɪ/ʃi:] と日本語の [ɕi]（し）の区別について，次のような対立を反復練習することを提案している．

(9) sea [sí:] — she [ʃí:] — しい（椎） [ɕí:]

また，歯音 [θ] は 3) 日本語にない音 であり，これを含む外来語（あるいは無理にカタカナに直した語）は「サンキュー」（< Thank you.）「シンク」（think）のように，[s] (または [ɕi]) で代用される．桑本 (2017:83) では，英語の [s] と [θ] の最小対立を練習する例を挙げている．

(10) sing [síŋ] 「歌う」 — thing [θíŋ]「もの」
sum [sʌ́m]「合計」—thumb [θʌ́m]「おやゆび」
mouse [máʊs]「ねずみ」— mouth [máʊθ]「口」

以上の [θ] [s] [ʃ] [ɕ] については，調音点を図示することは微細な差異を認識する上で極めて有効であると思われる（図 4）.

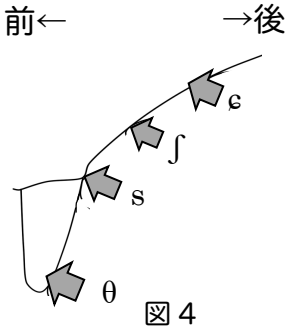


図４

桑本 (2017:84)，桑本・中村 (2017)

４．まとめ

以上，桑本 (2017) による子音の効果的な教授法について述べた.

桑本 (2017) は，英語の子音を調音点，調音様式によらず，日本語との異同や類推のしやすさによって 3 つに分類し，そのそれぞれについて順を追って説明することが，日本語から英語の音韻体系へ意識的に移行できる，あるいは発音実践をしやすいことを示した．本発表では，最も興味深い，また重要性が高いと思われる指導法として，3 項目について教授法を考察した.

発音指導の際には，共通して，以下の 3 点を意識することが特に重要だと思われる.

i. 日本語の音韻との異同を常に意識させる.

ii. 特にこれまで未知の音韻対立に対して明確に聞き分け，発音し分けを意識的に行わせる.

iii. 発音の実践をできるだけ個別に，丁寧に行う.

参考文献

桑本裕二 (2012)「発音指導に主眼をおいた「英語 LL 演習」の実践」『秋田工業高等専門学校研究紀要』第 47 号，99-105.

桑本裕二 (2017)『改訂版　小学校英語　発音のフシギ』秋田：秋田魁新報社.

桑本裕二・中村弘子 (2017)「英語授業における発音を見直そう！」平成 29 年度教員免許状更新講習ハンドアウト，2017 年 8 月 4 日，公立鳥取環境大学.

中村良夫・高橋邦行・Alexander McAulay・桑本裕二 (2015)『小学校英語の発音と指導—iPad アプリ「白柴さくらのえいごカルタ」読本—』東京：開拓社.

Haptic Techniques for Teaching English Prosody

Brian Teaman (Osaka Jogakuin University), Noriko Yamane (Hiroshima University)
teaman@wilmina.ac.jp, yamanen@hiroshima-u.ac.jp

1. Introduction

In this paper, we will focus on three prosodic variables of English: Intonation, stress and syllables. English and Japanese differ in many interesting ways, and do not allow for a one to one comparison of variables. For example, pitch in English is mostly a property of the intonational phrase whereas in Japanese, it is a property of words as well as the intonational phrase. The intonational phrase in English is linked to the sentence through the word that is in focus. Generally, though not always, this focal stress tends to be on the final content word in a sentence. In English there is word stress and sentence stress whereas in Japanese it is pitch accent (Beckman, 1986). In English accent there are three variables (pitch, duration and loudness) whereas Japanese has one chief variable--pitch. In English, duration is the primary marker of stress. Stressed syllables have a longer duration, which gives English its characteristic rhythm (see Adam & Munro, 1978 and Tajima et al., 1997). This variable, perhaps more than any other one variable contributes greatly to intelligibility (Munro & Derwing, 1995 and Anderson-Hsieh et al., 1992). Finally, Japanese syllables are mainly open as in V and CV, whereas English also has closed syllables CVC as well as consonant clusters consisting of 2 to 4 consonants. Transferring Japanese syllable structure to English greatly hinders intelligibility so getting syllables right or closer to English norms gives a boost to intelligibility.

Because of these substantial differences between Japanese and English, we suggest a haptic approach, which means using body movement and touch, as a promising way to teach and learn these prosodic aspects of English. After a brief look at some anecdotal evidence and previous studies, we will give a brief introduction to haptic techniques and then show the results of a pilot study.

2. Anecdotal evidence: One teacher’s experience

As a teacher of Japanese students for more than 20 years, I have observed that there is a barrier between studying English as a primarily internal, silent and cognitive activity to one which engages the entire body. Speaking is a physical act and involves the control of parts of the body much like a sport or dance. One can study about how to play tennis, but not achieve any proficiency without actually doing it and automatizing the many processes involved in playing a successful tennis match. The first days of university classes show that most students are very uncomfortable with starting to speak. It shows that they are getting very little practice with the physical part of English. In workshops for high school students, it is hard to get students to go beyond just siting and listening to the teacher. Haptic practice serves as a way to break out of the invisible shell that surrounds them and begin moving and speaking. Anyone can tell by the way that students look around nervously when you ask them to stand up, move their hands and arms, or even speak, that they are not used to crossing this line in the classroom. However, once they stand up they will then be more willing to

speak or perform haptic activities. This can be the first step to engaging the body and becoming a successful English speaker.

3. Previous Studies: Haptic Stimuli

Body motion coordinated with the focus words and stressed syllables will help learners enhance their acoustic prominence, and make their L2 English more intelligible. Students can start by imitating simple beat gestures. McCafferty (2006) found that beat gestures such as a sharp up-and-down movement of the hand improves L2 learners' syllable parsing. It is argued this metaphoric gesture helps L2 learners to gain control over syllable parsing, and the physicalized sense of prosodic features of the L2 improves automaticity and fluency. Krahmer and Swerts (2007) tested how acoustic prominence is perceived by observers when they *see* a speaker perform a manual beat gesture on a word, and found that they are perceived as more prominent than when they do not see the beat gesture. They also tested whether producing a manual beat gesture, a head nod and a rapid eyebrow movement lead to changes in acoustic prominence in speech, and found a significant effect with regard to duration (longer), energy (more intense) and peak *F*0 (higher).

Parrell, Goldstein and Byrd (2014) argue that the implementation of prosody is not domain-specific but relies on general aspects of the motor system. In their study, participants were instructed to tap their fingers and produce repetitive monosyllabic words simultaneously, and enhance one of them while keeping the other action unchanged. In spite of the instruction, when the finger tapping is enhanced, the oral articulator is enhanced, and vice versa. The results suggest speech and manual motor systems tend to coordinate with each other spontaneously.

4. Implementation of Haptic Techniques - Pilot

In this presentation, haptic techniques (Acton, 2016) will be introduced that have been shown to be helpful in the learning of these important prosodic features based on experience with learners from many contexts around world and including high school and first year college students in Japan. Haptic phonetics is a kinesthetic learning of sounds with the use of body movement and touch. Haptic learning helps make prosodic features salient and learnable (Acton et al., 2013, Teaman & Acton, 2013). Haptic movements perform two functions. One as a symbolic representation of an important sound feature and the other is a way to connect the body and performance of language to that symbolic representation. The following describes three haptic techniques used for intonation, rhythm, and syllables.

The Haptic Intonational Touchinami: The Haptic Intonational Touchinami is a blend of the words "touch" and the Japanese word for wave. It symbolically represents the shape of the melody, like a wave, and the intonational focus of the sentence. The intonational contour is represented by the basic shape of the contour of one hand moving across the visual field in a way that roughly matches the pitch pattern of the sentence. On the center of the intonational phrase, the wave hand strikes the other hand to symbolize the IP nucleus. (The application of this technique to intonation in a narrative is discussed in Yamane & Teaman (2017)).

The Haptic Rhythm Fight Club: In the haptic rhythm fight club, has the learner with arms raised in something like a boxing stance, with fists clenched. As the learner speaks, one hand thrusts forward on stressed syllables with either a long stroke forwards for main stresses or a short stroke forward for secondarily stressed syllables. This activity helps learners to feel the rhythm of English. It is especially effective in learning to engage the diaphragm and other core muscles used to expel breath with force on stressed syllables. (The application of this technique to focus words is discussed in Yamane & Teaman in prep, see Burri et al. 2016 for vocabulary learning.)

The Haptic Syllable Butterfly: The syllable butterfly helps students practice syllable structures and stress. With one hand cupping the opposite elbow and the other hand placed on the opposite shoulder, they tap harder on the shoulder (**TAP**) for the main stressed syllable and softer on the elbow (**tap**) for non-main stressed syllables. So the word "economy" is said with a **tap TAP tap tap** accompanying the 4 syllables of this word. This is an excellent teaching tool, so that when a student makes a stress error, the teacher can tap out the rhythm in front of the student and they can learn to correct the error themselves by applying the tapping pattern to their pronunciation. (The application of this technique to English rhythm is discussed in Burri & Baker 2016.)

5. Acoustic Analysis and Auditory Judgements

This section shows our approach to assessment, with preliminary results from a pilot haptic study that we conducted with university students (Yamane, Teaman & Acton, 2017). The average English proficiency level was 'pre-intermediate level' to 'intermediate level' (embassyenglish.com). A comparison group (n=10) received shadowing, while the experimental group (n=10) received haptic treatment (Intonational Touchnami). Both treatments were less than 15 minutes. Their task was to read an English passage in English textbook used in middle school (Kairyudo 2014). The target was "Everyone in her class loved the colorful flowers" located in the last sentence of a short narrative passage. The duration ratio of the stressed vowel "l[o]ved" in relation to "th[e]" (Ratio 1) and the duration ratio of the stressed vowel "cl[a]ss" in relation to "th[e]" (Ratio 2) were measured.

Results of the dependent (paired) sample t-tests indicated that there were significant differences in these ratios between pretest and posttest: [1] Audio group: (i) Ratio 1 between pretest (M=1.73, SD=0.92) and posttest (M=3.67, SD=2.73) conditions; $t(18)=2.13$, $p<.05$., and in (ii) Ratio 2 between pretest (M=2.31, SD=1.46) and posttest (M=4.21, SD=2.71) conditions; $t(18)=1.95$, $p<.05$. [2] Haptic group: (i) Ratio 1 between pretest (M=2.54, SD=1.59) and posttest (M=4.92, SD=2.01) conditions; $t(18)=2.93$, $p<.01$, and in (ii) Ratio 2 between pretest (M=2.51, SD=1.33) and posttest (M=4.91, SD=1.83) conditions; $t(18)=3.37$, $p<.01$. These results suggest that both kinds of practice have a positive effect on students' performance of making the contrast between stressed and unstressed syllables. Thus, their M values in the posttest approached those of native speakers' (n=10) (M=5.79, SD=3.48 for Ratio 1, M=4.99, SD=1.83 for Ratio2). However, there was no significant difference in the effects between the two groups: (i) Ratio 1; $t(18)=0.39$, $p=0.35$ (ii) Ratio 2; $t(18)=0.67$, $p=0.26$. Furthermore, one native speaker's blind auditory judgement (5 level likert-scale)

revealed that both groups show some improvement, but the pretest vs. posttest differences were not significant for the audio group (p=0.21) or the haptic group (p=0.14).

6. Conclusion and Further Direction

It can be seen that even though Japanese university students seem to readily learn to recognize and predict focus (Yoshimura et al. 2015), their spoken control of prosodic variables in English needs to be improved including stress, rhythm and syllable structure. Haptic methods seem to offer a powerful tool for raising awareness as well as improving performance of these variables.

Since pedagogical research for prosody is quite new, issues to tackle are many. The results are promising but lead us consider that we need a more systematic and automated assessment, more carefully-developed tokens, longer training, and more raters. Oral readings for pre and post tests are commonly used, but the next question is whether the learners are able to apply this to other novel sentences, especially in spontaneous speech.

Teacher education should also be enhanced. The amount of training teachers have received in pronunciation pedagogy strongly affects their knowledge and their confidence in that area (Baker 2014). Baker also shows that one's own experience of L2 learning influences their belief as well. Thus in phonetics classes in Japanese universities, it is important for prospective teachers to learn to self-monitor, self-analyze and self-evaluate their L2 speech performance, to experience different teaching methods that they can readily adopt, and to grow their metalinguistic and metacognitive ability, which helps them to further create their own methods and/or materials.

Selected References

Acton, W. (2016). HICPR: Haptic-integrated clinical pronunciation research and teaching. Unpublished manuscript.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *Tesol Quarterly*, *48*(1), 136-163.

Beckman, M. (1986). Stress and non-stress accent. Dordrecht: Foris Publications.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, *45*(1), 73-97.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, *25*(1), 1-24.

Teaman, B., & Acton, W. (2013). Haptic (movement and touch for better) pronunciation. In N. Sonda & A. Krause (Eds.), *JALT 2012 Conference Proceedings* (pp. 402–409). Tokyo, Japan: JALT.

Yamane, N., Teaman, B. & Acton, W. (2017, March) Teaching Focal Stress and Rhythm: Auditory vs. Haptic Techniques. Poster presented at the annual International Convention and English Language Expo, Seattle, WA.

日本音声学会

The Phonetic Society of Japan

2017 年（平成 29）年度

第 31 回　全国大会　予稿集

2017（平成 29）年度

日本音声学会第 31 回全国大会　予稿集
Proceedings of the 31st General Meeting of the PSJ

2017（平成 29）年 9 月 30 日発行

編集・発行：日本音声学会

代表者：今泉　敏