A1

日本語母語話者の英語韻脚におけるリズム制御

小西隆之・近藤眞理子 (早稲田大学)
tkonishi@aoni.waseda.jp

1. はじめに

1.1. 学習者英語の韻脚リズム

第二言語としての英語(L2 英語)におけるリズムの習得は、円滑なコミュニケーションのために不可欠であることが先行研究により示されている(Derwing & Munro 1997, Saito et al. 2015 等)。昨今では英語教育においても学習初期段階における韻律指導の重要性が認知されはじめ、しばしば分節音指導に先んじて行われることもある。一方で L2 英語のリズム、特に英語のリズムにおいて重要な単位である韻脚(フット)に関する研究は未だ少ない。

1.2. 日本語母語話者の英語韻脚リズム

英語はストレスアクセント言語で(Beckman 1986)、そのリズムはフットに制御されると言われている(Féry 2018 等)。フットの大きな特徴として、その時間長がフット内の音節数(母音数)に関わらずに一定であるという等時性(isochrony)が挙げられる(Féry 2018, Gussenhoven 2004 等)。このフットの等時性は主に心理的なもので、Cruttenden (2014)などは英語のフットにおける物理的な等時性の存在を否定している。一方で、Lehiste (1975)や Tajima (1998)など、限定環境下における物理的等時性の存在を示す研究もある。

これに対して、日本語はそのリズムをモーラ(拍)に制御され、発話時間長が概ねモーラ数に比例する等拍リズムを有する(Port et al. 1987 等)ことが一般に知られている。つまり、発話の時間長が母音数に比例する。従って、日本語母語話者の英語韻脚リズムの実現においては母語のモーラリズムの強い干渉が想定される。

著者らは、L2 英語音声コーパス J-AESOP (Kondo et al. 2015 等)を用いて、日本語母語話者群 183 名(上級話者群 90 名、初級話者群 93 名)と英語母語話者群 25 名による The North Wind and the Sun の読み上げ文を分析した(Konishi 2019)。その結果、初級話者群の発話においては 1 音節フットの生起頻度が最も高く、フット内の音節数が増えるにつれて生起頻度が低くなるのに対し、母語話者群と上級話者群の発話においては 3 音節フットの生起頻度が最も高く、次いで 2 音節フットの生起頻度が高いことを明らかにした(図 1)。一方で時間長に関しては、フット内の音節数に比例して時間長が増加しており(図 2)、母語話者英語および L2 英語のフットに関して Cruttenden (2014)などが主張する物理的等時性の不在を支持する結果となった。

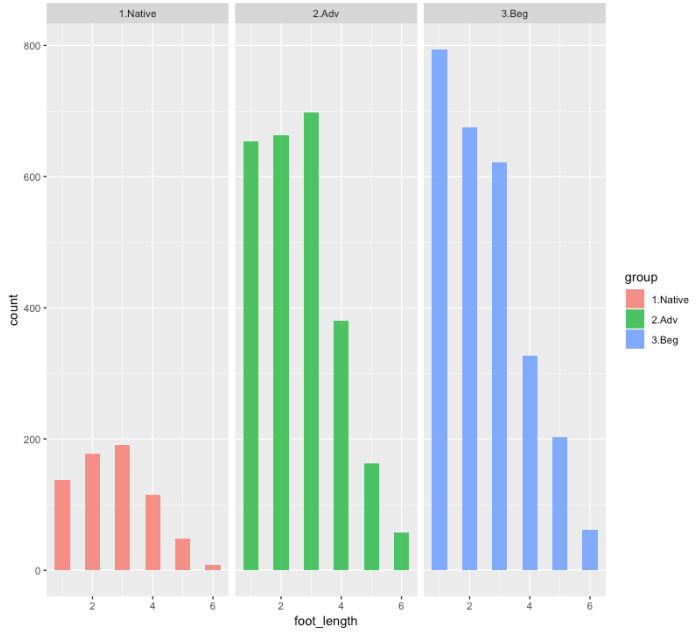


図 1: フット内の音節数と生起頻度

X 軸: フット内の音節数　Y 軸: 生起頻度

赤: 英語母語話者　　緑: 上級学習者　　青: 初級学習者

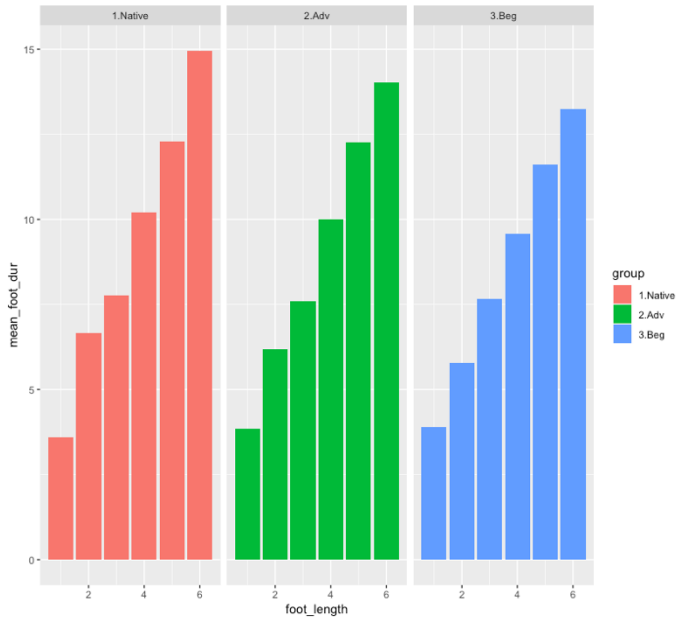


図 2: フット内の音節数と時間長

X 軸: フット内の音節数　Y 軸: 時間長

赤: 英語母語話者　　緑: 上級学習者　　青: 初級学習者

1.3. 本研究の目的

本研究では、Konishi (2019)の母語話者と上級話者の発話において生起頻度が最も高かった 2 音節フットと 3 音節フットに着目し、これらの実現においてどの程度の等時性の傾向が見られるかを、同一のデータを用いて各群で比較検証した。

2. 2 音節フットと 3 音節フットの時間長の分析

2.1. 音声データ

データには、Konishi (2019)と同様の L2 英語音声コーパス J-AESOP コーパスの中の *The North Wind and the Sun* の読み上げ文(以下)を用いた。

> The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew, the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shone out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

2.2. 学習者群の分類

Konishi (2019)と同様に、被験者群を英語母語話者群(25 名)、上級話者群(90 名)、初級話者(93 名)の 3 群に分けて比較を行った。日本語母語話者の分類の歳には J-AESOP に付与されている韻律評定値[1]を用いた。評定項目は 1)語彙的ストレスの実現、2)リズム、3)セグメントの誤挿入や脱落の 3 項目だった。

2.3. コーディングと分析

J-AESOP に付与されているのは分節音のアノテーションのみなので、分析のために韻律情報を機械的に取得した。本研究では、Cruttenden (2014)等に倣い、ストレスを担っている音節(強勢音節)から次の強勢音節の直前の音節までを一つのフットと定義した。一般に、英語には強弱型(trochaic)と弱強型(iambic)の 2 種類のフットが存在するが、今回は強弱型フットを前提とした計算を行った。この根拠として、イギリス英語コーパスを用いて、最頻語 13,000 語のうちの内容語の 73%が強弱型アクセントを持っていたとする Cutler (1989)の研究結果と、*lettuce － let us* や *invest － in vest* などの対を用いた英語母語話者に対する知覚実験の結果、強弱型アクセントの場合には 1 語、弱強型アクセントの場合には 2 語に知覚される傾向があることを示した Taft (1984)の実験結果などが挙げられる。これらを踏まえ、Tajima (1998)は、英語の語彙システムにおいて語頭ストレス(即ち強弱型アクセント)を好む心理的傾向があると主張している。

ストレスの音響パラメータとしては、母音の音圧、基底周波数(F0)、時間長、フォルマ

[1] 評定値に関する詳細は小西・近藤(2017)等を参照

ントの質の 4 つが一般に知られている。本研究に際し、この 4 つのパラメータ各々を用いてストレスを推計するプログラムを構築した[2]が、時間長を用いた推計の精度が最も高かった。

従って、まず母音の時間長を用いてストレスを機械的に同定した。話速に関して正規化するために、各母音の時間長を前後 5 語(当該の母音を含む語 + 前後 2 語ずつ)の時間長を当該 5 語の母音数[3]で割ったものを計算に用いた。そして話者ごとの母音の時間長の平均値を求め、母音の時間長が平均値の 2 倍の 60% (即ち平均値の 120%)を超えた場合に強勢音節と認定した。

次に同定したストレスを用いてフットを定義した。強勢音節から次の強勢音節またはポーズ(破裂音等の閉鎖区間以外の空白)の前の音節までを 1 フットと定義した。フット内の最大音節数を 6 とし、それを超えた場合には直前の最大時間長を持つ音節を強勢音節として再認定し、新たにフットを定義した。

最後に各々の話者について話速と Z スコアで正規化した母音の時間長を用い、まず、各話者に関して 2 音節フット、3 音節フットごとに時間長の平均値を算出し、それらを各群(母語話者群、上級話者群、初級話者群)間で比較した。比較の際にはデータとしたフット全てを音節数に関わらずに一律に扱った。

3. 分析結果と考察

分析の結果、標準化された時間長の分散のレンジが、初級話者群(2.83) > 上級話者群(1.46) > 母語話者群(1.11)の順に大きく、また、標準偏差(SD)も初級話者群(0.42) > 上級話者群(0.32) > 母語話者群(0.23)の順に大きいことがわかった(表 1)。従って、母語話者が最も等時性リズムを実現しており、学習者においては習熟度の上昇とともに韻脚リズムが等時性を持つように変化していくことが示された。本研究結果は、Lehiste (1975)や Tajima (1998)などによって示されていた、限定環境下における英語韻脚リズムの等時性の存在に新たな一例を加え、されにそれを L2 英語において再現した点で意義深いと言える。

表 1: 各群の 2,3 音節フットの時間長の分散のレンジと標準偏差

	レンジ	標準偏差(SD)
母語話者群	1.11	0.23
上級話者群	1.46	0.32
初級話者群	2.83	0.42

[2] 例えば、「音圧が平均値の 2 倍以上であれば強勢音節と認定する」など
[3] 日本語母語話者の英語には挿入母音があるため、音節数ではなく母音数を計算に用いた

4. 謝辞

本研究は科研費若手研究 B(17K13513)および基盤研究 B (15H02729)の助成を受けている。

参考文献

Beckman, M. E. (1986). Stress and Non-Stress Accent. In Van den Broecke, M.P.R and van Heuven., VJ (Eds). Netherlands Phonetic Archives; 7. Foris publications, USA.

Cruttenden, A. (2014). *Gimson's pronunciation of English*. Routledge.

Cutler, A. (1989). Auditory lexical access: where do we start? In *Lexical representation and process* (pp. 342-356). MIT Press.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in second language acquisition*, *19*(1), 1-16.

Féry, C. (2018). *Intonation and Prosodic Structure.* Cambridge University Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Kondo, M., Tsubaki, H., & Sagisaka, Y. (2015). Segmental variation of Japanese speakers' English: Analysis of “the North Wind and the Sun” in AESOP Corpus. 音声研究 *19*(1), 3-17.

Konishi, T. (2019). Acquisition of L2 English Foot Rhythm by Native Speakers of Japanese. New Sounds 2019. Aug 30-Sep 1. Waseda University. Japan.

Lehiste (1975). The perception of duration within sequences of four intervals. In *Proc of ICPhS VIII.* Leeds, UK.

Port, R. F., Dalby, J., & O’Dell, M. (1987). Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America*, *81*(5), 1574-1585.

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37*(2), 217-240.

Taft, L. (1984). *Prosodic Constraints and Lexical Parsing Strategies.* Doctoral dissertation, University of Massachusetts.

Tajima, K. (1998). *Speech Rhythm in English and Japanese: Experiments in Speech Cycling.* Doctoral Dissertation, Indiana University.

小西隆之、近藤眞理子(2017)「L2 英語音声の評定における評定者の母語の影響 (Cross-linguistics Influence on L2 English Speech Rating)」第 31 回日本音声学会全国大会予稿集 pp138-141.

A2

日本語での中和環境における英語音/f/, /h/の知覚 ー後続母音との cross-splicing による検証

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

本研究では，日本語の音韻規則（fu-hu の中和）が引き起こす英語音/ f /と/ h /の同定困難（青柳・Wang, 2017）について，その要因が子音部の音響特徴ではなく，後続する/ u /の提示そのものにあることを，子音部と各種後続母音との cross-splicing によって検証した．これによって，母語の音韻規則の影響を改めて示すと同時に，音声知覚が音節単位でなされており，母音を聞き終わった後に逆行して子音を同定し直すプロセスの存在を検討する．

英語の/ f /と/ h /はそれぞれ日本語の/ ɸ /と/ h /として知覚同化され，/ 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]となるため, / ɸu /と/ hu /はともに[ɸu]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*[1]（/ fɑ /-/ hɑ /）の同定は容易である一方，*Fu-who*[2]（/ fu / - / hu /）の同定は困難となり得る．

青柳・Wang（2017）は，*fee/he, foe/hoe, Fu/who* 等の有意味語，およびそれらの語から抽出した子音部を用いて同定実験を行った結果，(i) *Fu/who, few/Hugh*[3] の識別感度（*d'*スコア）[4] が著しく低く同定が困難であること， さらに， (ii) /fɑ, hɑ/や/fi, hi/から抽出した摩擦部/ f /, / h /のみの同定はほぼ問題がなく，抽出元全体の/fɑ, hɑ/, /fi, hi/では同定が向上する一方で，/fu, hu/から抽出した/ f /, / h /はより難しくなるのみならず，抽出元全体の/fu, hu/を提示すると，子音部のみよりもさらに感度が低下することが示され，/ u /の提示がもたらす母語の音韻規則（fu-hu の中和）の知識が強く影響していることが示唆された（2.3 節参照）．

しかし，各母音環境から抽出された/ f /, / h /はその後続母音の同時調音情報を含むので，上記の結果は，中和規則の影響だけでなく，/fu, hu/から抽出した子音（ f(u), h(u)）が他の母音環境（ f(a, i), h(a, i)）よりも音響的に類似していることにより同定が困難になったという可能性を排除できない．そこで本研究では，/a, i, u/環境（*fa/hah, fee/he, Fu/who*）から抽出した/ f /と/ h /（ f(a)/h(a), f(i)/h(i), f(u)/h(u)）と後続母音（ _a, _i, _u）を cross-splice した刺激を用いて同定実験を行い，/ u /の環境の影響をさらに検証した。

2. 先行研究

2.1. / f /と/ ɸ /

日本語にはない無声唇歯摩擦音/ f /は，一般に無声両唇摩擦音/ ɸ /に知覚同化され，英語の/ f /-/ h /の対立は日本語の/ ɸ /-/ h /の対立として概して保たれる（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その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辺りでも出現する．この

[1] *fa* は音符名（ドレミ「ファ」），*hah* は間投詞．[2] *Fu* は中国人名「溥」「付」（*Mr. Fu*）．[3] *Hugh* は人名「ヒュー」．
[4] 強制二者択一法では一刺激ごとの正答率は同定力を反映し得ないため，本稿でも，信号検出理論に基づいて f と h への反応の間にある偏りから算出する *d'*を使用し，f/h ペアとしての同定識別力を検討する．

ように，日本語において調音的・聴覚的に区別する必要がないことから，これが転移して英語でも/fu/と/hu/の同定は難しくなり，これは/fa/と/ha/等と比べても明らかである．

外来語借入の古いものは，/ f /は母音つきの/ɸu/で代用された（Irwin, 2011）（例：*fan* > /ɸu.an/ “フアン”，*film* > /ɸu.i.ru.mu/ “フイルム”，*felt* > /ɸu.e.ru.to/ “フエルト”）．後に，/uV/が二重母音化，さらには/ u /が脱落して，今日では/ f /は子音/ ɸ /単体で代用されることが多い（例: *fan* > /ɸan/ “ファン”，*film* > /ɸi.ru.mu/ “フィルム”，*felt* > /ɸe.ru.to/ “フェルト”）．しかし，これら新旧の音形は必ずしも二分別的ではない．二音節書きと一音節書きの混在は未だ多く，また音声的にも，“フアン”と“ファン”等では [ɸɯ.aɴ] ⇔ [ɸ^{ɯ}an] ⇔ [ɸan]のような連続性はよく聞かれることからも，/ f /は/ ɸ /単体で代用されているとはいえ，/fV/を/ɸɯV/ととらえること（例: /fa/を/ɸɯa/, /fi/を/ɸɯi/ 等）は整合性がある（3.2.2 節参照）．

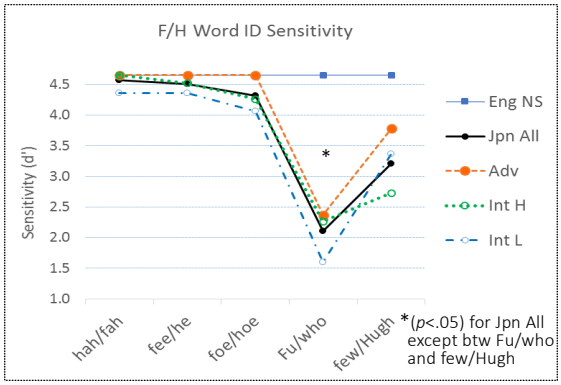
2.2. / f /と/ h /の知覚

英語摩擦音の音響的な特徴とそれに基づく知覚については Jongman *et al.*（2000）等が詳しいが，/ h /は含まれていない．/ h /は特定の狭窄点を持たず固有の音響特徴が少ないことから，摩擦音の音響的・知覚的比較に含まれないことが多い．

また，日本語話者による/fu/, /hu/の知覚については報告が多くない．Guion, *et al.* (2000) は日本語話者による英語子音知覚を詳細に検討しているが，困難な英語子音対（l/r, th/s, v/b 等）の中に f/h を含んでいない．数少ない実証実験の一つとしては Lambacher *et al.* (2001) がある．日本人大学生 100 余名を対象に無意味音節を使った英語子音同定実験を行った結果，/ f /と/ h /は，/ u /の環境（/_u/，/u_u/，/u_/）で正答率と識別感度が低く，クラスター分析によって，/ u /の環境の/ f /と/ h /が最も近く分類されると示しているが，この識別困難の要因については考察されていない．

2.3. 青柳・Wang（2017）— *Who is Fu?*

青柳・Wang (2017) は，英語話者と日本語話者を対象として，*fee/he, fa/fah, foe/hoe, Fu/who, few/Hugh*，および *fee/he, fa/hah, Fu/who* から抽出した子音部 f(i)/h(i), f(a)/h(a), f(u)/h(u)を刺激として，語と子音の同定実験を行った．図 1 は各対語の識別感度（*d’*スコア）である．



主な結果として：

- 英語話者（Eng NS）は母音環境にかかわらず，f//h 識別に問題は無い．
- 日本語話者 (Jpn)は/ u /の環境（*Fu/who, few/Hugh*）で識別感度が低く，同定が困難．
- 学習上級者でも，/ u /環境で f/h の同定が困難．
- 語彙効果（頻出語に回答する偏り）がある．（*Fu* より *who*，Hugh より *few* を回答）

図 1．f/h 語彙同定の識別感受性（EngNS，Jpn）（青柳・Wang，2017：図 2 より）

次に，表 1（抽出子音部のみと，その抽出元の語全体の識別感度の比較）を見ると，日本語話者も各語から抽出した子音 f/h のみでの同定ができることがわかる，特に f/h(a)と f/h(i)ではほぼ問題ない（f/h(a)：*d’* = 4.37．f/h(i)：*d’* = 4.05 等）．また，f/h(u)では感度は下

表 1. 「抽出された C」と「抽出元の語全体 CV」の識別感度比較（Jpn）

	抽出 C f/h(a)		CV fah/hah	抽出 C f/h(i)		CV fee /he		抽出 C f/h(u)		CV Fu/who
Jpn All	4.37	↗	4.57	4.05	↗	4.51		2.91	↘	2.10
Advanced	4.65	↗	4.65	4.42	↗	4.65	VS	3.89	↘	2.37
Int'med H	4.50	↗	4.65	3.64	↗	4.52		2.67	↘	2.26
Int'med L	3.89	↗	4.36	4.31	↗	4.36		2.31	↘	1.60

がるものの，*d'*は 3 前後で，ある程度は識別できると思われる．

注目すべきは，*Fu/who* から抽出した f/h(u)の感度は（*d'* = 2.91 等），その抽出元の全体である *Fu/who*（*d'* = 2.10 等）よりも高いことである．通常，C 単独よりも CV の方が情報量も多く C の同定が容易になると思われる（C < CV）．実際，英語話者の全刺激と日本語話者の f/h(a, i)ではそのような反応であった．一方，/ u /環境では，単独で聞こえていた C が，その C から始まる CV 全体が提示されると聞こえなくなるという逆転現象が起こっている．/ u /を聞くと，日本語では音素対立環境ではなく異音変異であるという音韻規則の知識とそれに基づく音声知覚の経験があることが要因であると示唆された．

しかし，各母音環境から抽出された子音 f/h は，後続母音の同時調音の情報が含まれており，/ u /の環境（語および抽出子音）で f/h の識別感度が低下するのは，この環境における f/h が他の母音環境でよりも音響的に近く，混同しやすかったからであるという可能性を捨てきれない．そこで，各母音環境から抽出した子音に，別の母音を挿げ替える cross-splicing を行い，その同定実験を行った．予測通りに，識別に問題のなかった f/h(a)や f/h(i)の子音単独に/ u /を接続した際に識別が低下し，f/h(u)の子音に/ a /や/ i /を接続した際に識別が向上するとすれば，/ u /環境における f/h の識別の困難は，その子音自体ではなく後続の/ u /の提示そのものにあるという仮説が裏付けられることになる．

3. 実験

3.1. 方法

＜刺激＞ 青柳・Wang (2017) で使用した *fee/he, fa/fah, Fu/who*（4 人の発話）から抽出した子音部: f(i), h(i), f(a), h(a), f(u), h(u)と，各語の母音部: _i, _a, _u をそれぞれ cross-splice した（接続した刺激は"f(i)_a"のように表記）．/f /には *f* の語の母音部を，また/ h /には *h* の語の母音部）を接続した．なお，f(i)_i 等，原音（6 語）に戻るような操作はしないこととする．

f/h(i)
f/h(a)
f/h(u)　子音 2 種（×抽出元母音 3 種）　×　_i
_a
_u　後続母音 3 種 = 12 種*　(18－6 = 12)

*原音 6 種を除く

また，比較のために青柳・Wang（2017）の同定実験を追試し，*fee/he, fa/hah, Fu/who* および抽出子音のみ f/h(i, a, u)も分析対象とした．よって，(1) 4 人の音源，(2) 24 刺激（(i)子音×母音の cross-splice 刺激 12 種, (ii) 上記の語全体 6 種, (iii) 抽出子音部のみ 6 種），(3) 4 回繰返し，合計 384 個の刺激が準備され，第一ブロック（上記(i)と(ii)の語彙同定）と第二ブロック（(iii)の子音のみの同定）に分けて音声分析編集ソフト *praat* の実験プログラム上に搭載された．第一ブロックは語の判断を求めるもので，これを先に実施することにより，第二ブロック実施によって起こり得る子音部のみに集中するという影響の軽減を図った．

＜被験者と手順＞ 実験には日本語母語話者大学生 20 名（19 - 22 歳）と参考比較群として英語母語話者 7 人（40 - 68 歳）が参加した．被験者は実験語の音と意味の確認をしてから，静かな部屋で実験プログラムによってヘッドセットからランダムに提示される音刺激を聞き，PC 画面上に表示される語または子音をキー操作よって強制二者択一した．フィードバックはなく，回答をすると約 2.5 秒後に次の刺激が提示された．最初に手順の練習をした後，休憩を 4 回入れながら，全体で 30 分前後の実施であった．

3. 2. 結果と考察

3. 2. 1. 追試結果（抽出子音 C と語全体 CV）

青柳・Wang (2017) の追実験の結果として，英語母語話者は語の同定は完璧で（*d'* = 4.65, 理論的最高値），子音のみもほぼ完璧であった（f/h(a): 4.65, f/h(i): 4.52, f(u): 4.42）．次に，日本語話者の語と抽出子音のみの識別感度を表 2 に示す．まず，右列が示すように，*fa/hah* および *fee/he* では同定に問題がなく（*d'* = 4.27, 4.53），一方で，*Fu/who* は著しく困難になる（*d'* = 1.70）ことが再確認された（対語種を要因とした ANOVA：$F(2, 38) = 33.125$, $p < .01$, Bonferroni で *Fu/who* に有意差: $p < .01$）．次に，左列が示すように，各語から抽出された子音のみでの同定が特に f/h(a)と f/h(i)で高く（*d'* = 4.25, 4.38），f/h(u)では下がるものの（*d'* = 3.14）ある程度は識別できていると思われる（抽出元母音環境を要因とした ANOVA：$F(2, 38) = 14.598$, $p < .01$, Bonferroni で f/h(u)に有意差: $p < .01$）．さらに重要なことに，子音部 f/h(u)での同定よりも，抽出元の語全体（*Fu/who*）の方が困難になることが追認された．

表 2. 「抽出子音部のみ」と「抽出元の語全体」の同定（識別感度）の比較（日本語話者）

Excised C only	*d'*		Whole word	*d'*
f/h (a)	4.26	↗	fa/hah	4.27
f/h (i)	4.38	↗	fee/he	4.53
f/h (u)	3.14	↘	Fu/who	1.70

3. 2. 2. 子音×母音の cross-splicing

表 3 は，抽出子音部と別の母音を cross-splice した刺激の語彙同定の結果である．英語話者は全ての組合せにおいて，また日本語話者も多くの組合せにおいて，抽出子音単体でよりも cross-splice 刺激で識別感度が下がっている．これは，相反する音声をつなぎ合わせた非自然音声であるので想定内であるが，よく見ると両者間で反応パターンが違っている．

表 3. 「子音×母音を cross-splice した刺激」の同定（a. 英語話者，b. 日本語話者）

Excised C only		spliced w/ _a	_i	_u
f/h(a)	4.65	—	4.42	3.63
f/h(i)	4.52	**2.68**	—	3.17
f/h(u)	4.42	3.65	4.25	—

(a) 英語話者

Excised C only		spliced w/ _a	_i	_u
f/h(a)	4.26	—	4.14	**1.83**
f/h(i)	4.38	4.08	—	**2.04**
f/h(u)	3.14	3.73	3.78	—

(b) 日本語話者

英語話者は，f/h(i)に_a と_u が後続した際に最大の低下を示し（f/h(i): 4.52 ↘ f/h(i)_a: 2.68），↘ f/h(i)_u: 3.17），日本語話者は，f/h(a), f/h(i)に_u が後続した際に顕著な低下を示している

（f/h(a): 4.26 ↘ f/h(a)_u: 1.83，および f/h(i): 4.38 ↘ f/h(i)_u: 2.04）．また，日本語話者は，子音単独よりも splice 刺激で感度が低下する全体的傾向の中で，f/h(u)に_a と_i が後続した際には上昇している（f/h(u): 3.14 ↗ f/h(u)_a: 3.73，および ↗ f/h(u)_i: 3.78）．この低下と上昇のどちらも予測通りの結果であるが，とりわけ，子音単独で同定に問題のなかった f/h(a, i)が，/u/が後続すると f/h の混同が起こってくるのは，まさに，子音ではなく母音/u/の影響であることを示す結果となった．

図 3 は，cross-splice 刺激の各組合せについて誤答を示したものである（例：f(a)に_u を接続したときに *Fu* ではなく *who* と回答）．まず，英語話者はどの/f/にどの母音が後続しようと/f/と聞き，/h/と聞くことはない（図 3 上，左側）．しかし，/h/に別母音が後続すると/f/と聞くケースが現れる（同右側）．h(i)に_a や_u，また h(u)に_a が後続すると，それぞれ *hah, who, hah* ではなく，*fa, Fu, fa* と反応することがかなりの率で見られる．この/f/と/h/の違いは，従前から言われる音響的な違いにあると思われる．/f/は特徴的なスペクトルを持ち，かつ唇の開きが後続母音開始時のフォルマント推移に現れ，この両者が/f/の知覚に寄与する（Jongman *et al.*, *ibid*）一方で，/h/のスペクトルピークは後続母音のそれを先行しているのでフラットなままでフォルマント推移がない（竹林, 1996）．/h/の独自性というよりも，いわば，後続母音の無声開始部である．

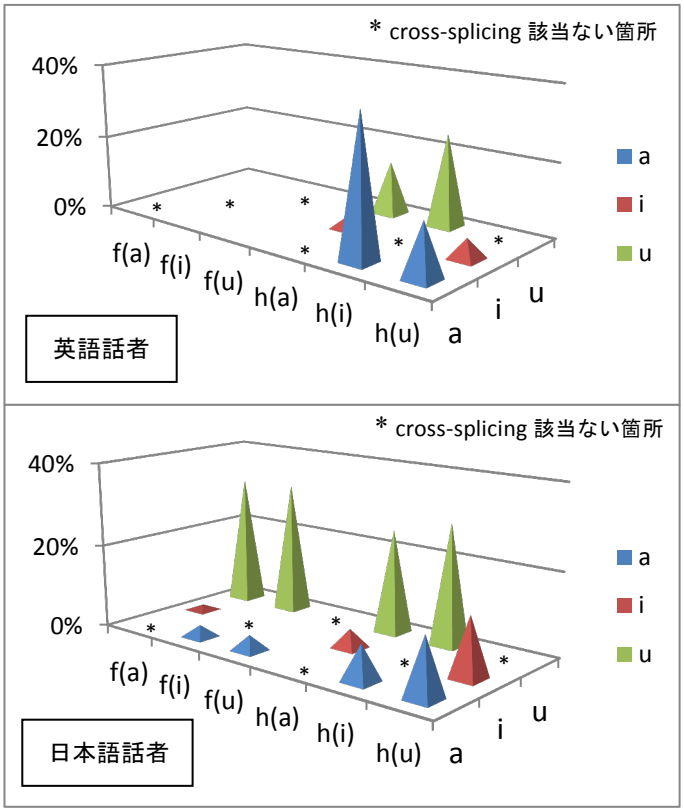


図３．Cross-splicing 組合せと誤答

実際，f(a, i, u)を聞いても抽出元の母音はほぼ分からないが，h(a, i, u)では分かることが多く，実験後調査でも多くの被験者がそう答えている．よって，最も狭い調音であり，/f/のスペクトルピークに近づく/i/の F2 の響きを持つ h(i)から，_a や_u に開いていくフォルマント推移を，/fV/のそれと聞くこともあると思われる．このように，抽出した/h/に別の母音を後続させると（例：h(a)_i），/h/の疑似フォルマントは母音フォルマント開始時に突如変化を起こすので自然度が落ちることがある．その結果，両方の母音が聞こえることもあり，反応のパターンも若干複雑なものとなった．その点，/f/の splice 刺激は自然で，母音は一つしか聞こえない．/f/の splice 刺激を中心に結果を解釈することが妥当であると思われる．すると，英語話者は，/f/にどの母音が後続しても/h/に聞くことがない（cf. Foulkes, 1997）ことが参照点となる．

一方，日本語話者が/f/を/f/と聞くのは後続が_a/と_i/のときのみであって，_u の環境ではかなりの率で/h/と混同する．f(a), f(i)は抽出子音単独で聞こえていたが（表 3(b), 左列），

_u が後続すると（f(a)_u, f(i)_u）困難になる（図 3 下，左奥）．/ u /の環境での識別困難はペアである/ h /側にも現れていて，単独で聞こえていた h(a), h(i)は，_u が後続するとかなりの率で/ f /と誤る（同，右奥）．英語話者の正誤が/ f /, / h /で異なることは音響的な特徴に起因していると思われるが，日本語話者の正誤は，/f, h/にかかわらず，また/f, h/が持つ抽出元母音からの同時調音情報にかかわらず，/ u /の提示そのものが影響していることがわかる．

さらに，上記とは別に，/hu/に関連して興味深い結果が得られた．日本語話者が h(u)に _a や_i が後続すると/ f /を聞くことである（図 3 下，右手前）．h(u)も抽出元母音の/ u /の音色を持っており，そこに_a や_i が後続すると [ɸɯa] ～ [ɸ^{ɯ}a] や [ɸɯi] ～ [ɸ^{ɯ}i] となり，これを/fa/や/fi/と聞いたのではないだろうか．/ f /を日本語化した際の特徴が影響した可能性が考えられる（2.1 参照）．音声の推移を聞くという点では一緒だが，英語話者（h(i)_a を/fa/と聞く）と日本語話者（h(u)_i を/fi/と聞く）ではその捉え方が違うことがわかる．

4. 結論と今後の展望

以上見てきたように，日本語話者は，子音単独では/ f /, / h /の同定がかなりできるが，/u / 環境の語（*Fu/who*）では困難となる．また，/ a /や/ i /の環境から抽出された（i.e. / u /の特徴を持たない）/f, h/は同定に問題がなかったものが，/ u /が後続すると困難になることから，その子音ではなく，/ u /の提示そのものが要因であることが裏付けられた．

/ u /の提示は，日本語では音素対立環境ではなく異音変異を示すもので，より高次の音韻規則の知識が，“今不要な”音声レベルの知覚能力を鈍化または潜在化させるプロセスが示唆される（cf. Boomershine *et al.* 2008）．また，一連の結果は音節単位での音声知覚に一致している．頭からセグメント毎に順次処理するとすれば，C のみので聞こえていたものが，その C から始まる CV で聞こえなくなることはない．冒頭から入っているであろう C の音響情報を，V まで聞いてその処理の仕方を変えている可能性を支持する結果となった．

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した．ここに感謝します．

引用文献

青柳 真紀子・Wang, Yue（2017）「日本語での中和環境における英語音の知覚 － *Who* is *Fu*?」『第 31 回日本音声学会全国大会予稿集』（pp. 142-147）

竹林 滋（1997）『英語音声学』東京：研究社．

Boomershine, A., Hall, K., Hume, E. & Johnson, K. (2008) “The impact of allophony vs contrast on speech perception,” In Avery, Dresher & Rice (eds) *Phonological Contrast*. N.Y.: Mouton de Gruyter.

Foulkes, P. (1997) “Historical laboratory phonology – Investigating /p/ > /f/ > /h/ changes,” *Language and Speech*, 40 (3), 249-276.

Guion, S., Flege, J., Akahane-Yamada, R. & Pruitt, J. (2000) “An investigation of current models of second language speech perception The case of Japanese adults’ perception of English consonants,” *J. Acoustical Society of America*, 107 (5), Pt. 1, 2711-2724.

Irwin, M. (2011) *Loanwords in Japanese*. Amsterdam: John Benjamins

Jongman, A., Wayland, R. & Wong, Serena. (2000) “Acoustic characteristics of English fricatives,” *J. of Acoustical Society of America* 108 (3), Pt. 1, 1252-1263.

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.

A3

Acquiring Jaw Movement Patterns in a second language: Some lexical factors

Ian Wilson (University of Aizu), Donna Erickson (Haskins Laboratories, Kanazawa Medical University), Shigeto Kawahara (Keio University), Tomoko Monou (International Christian University)
wilson@u-aizu.ac.jp, ericksondonna2000@gmail.com,
kawahara@icl.keio.ac.jp, nndbt653@yahoo.co.jp

1. Introduction

Understanding how we acquire the production patterns of a second language (L2) is an important topic in phonetic studies. At least among those studies conducted in Japan, it is probably safe to say that research on how Japanese speakers acquire English is far more common than how English speakers acquire Japanese. Among the latter type of studies, common topics include the acquisition of Japanese pitch accent and long-short vowel/consonant distinctions, (e.g. Shport 2016, Hirata et al. 2007). Here, we examine a yet little explored aspect of L2 prosody: Japanese phrasal stress, as manifested via systematic jaw movement patterns, and the articulation thereof by English speakers of Japanese. As background, we are working within a theoretical framework (e.g. the C/D model: Fujimura 1992) in which there are at least two "prosodic articulators": the larynx, which controls pitch (F0) as well as vowel duration and intensity; and the jaw, which controls changes of the resonant frequencies of the vocal tract. Previous studies have observed that increased jaw displacement results in more prominent syllables (e.g. Fujimura 1992) as well as higher F1 values (e.g., Erickson 2002, Erickson et al. 2012). Strong correlations between measures of jaw displacement, F1 and n-ary levels of syllable prominence for each syllable in the utterance have been reported for English (Erickson et al 2015). For languages such as Chinese, French, and Japanese, increased jaw displacement/F1 associates not with n-ary levels of syllable stress, but with phrase final words/syllables (Smith et al. forthcoming, Erickson et al. 2016, Kawahara et al. 2014, 2015). Generally speaking, first-language (L1) prosody can be carried over to the L2, and jaw movement patterns are not exceptional. For instance, Japanese speakers of English show increased jaw displacement/F1 phrase finally, even in sentences where L1 English speakers have reduced jaw displacement/F1 in that position (Erickson et al. 2014). The question we address here is what are the patterns of jaw displacement/F1 of English speakers speaking Japanese as an L2 in comparison with those of L1 Japanese speakers.

2. Method

Articulatory and acoustic recordings were made using 3-D EMA (Carstens AG500 Electromagnetic Articulograph) at the Japan Advanced Institute of Science and Technology (JAIST). One sensor was placed on the lower medial incisors to track jaw motion. Additional sensors were placed on three points of the tongue, and the upper and lower lips, but the analyses of these sensors are not reported.

Four additional sensors (upper incisors, bridge of the nose, left and right mastoid processes behind the ears) were used as references to correct for head movement. The articulatory and acoustic data were digitized at sampling rates of 200 Hz and 16 kHz, respectively. The occlusal plane was estimated using a biteplate with three additional sensors. In post processing, the articulatory data were rotated to the occlusal plane and corrected for head movement using the reference sensors after low-pass filtering at 20 Hz. Custom software – Mview (Mark Tiede, Haskins Laboratories) was used to analyze the data. The lowest vertical position (maximum displacement) of the jaw with respect to the bite plane was located for each target syllable of the utterance using a velocity-based criterion. F1 measurements in the mid region of the vowel were made using Praat software.

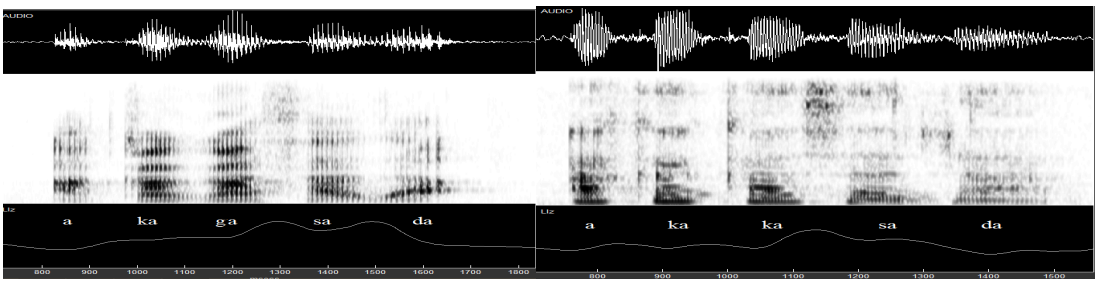


Figure 1: Acoustic waveform, spectrogram, jaw position for [akagasada] for Japanese (left) and American English (right) speaker. L2 speakers said [aka kasa da] failing to apply rendaku.

The L1 speakers spoke standard Tokyo Japanese (1 male, 2 females, in their 30s). The L2 speakers were 3 North American English speakers (2 males, 1 female, approximate ages 30 to 60), fluent in Japanese, having lived in Japan for the past approximately 8 to 17 years. The stimuli, together with stimuli for other experiments, were presented sentence-by-sentence in a randomized order on a PowerPoint screen in front of the speaker. Each speaker pronounced each stimulus approximately 6 times, but due to tracking errors or mispronunciation errors, sometimes the number of tokens varied. Since the magnitude of jaw displacement is nontrivially affected by vowel height (Kawahara et al. 2014, Menezes and Erickson 2013), all vowels of the stimulus sentences were [a]. The sentences were *aka-gasa da* (“That’s a red umbrella.”) and *aka-pa'jama da* (“Those are red pajamas.”). Previous analysis of some of this data, focusing on the production of L1 speakers, was reported in Kawahara et al. 2014; some data are reproduced here for the sake of comparing L1 and L2 speech.

3. Results

3.1. L1 and L2 patterns of jaw displacement and F1 for *aka kasa da*

The articulatory results in the following figures show the amount of average jaw displacement for each syllable as measured from the occlusal plane across the 5 to 6 repetitions of the utterance. The height of each bar indicates the amount of jaw displacement: the higher the bar, the greater the jaw

opening, (i.e., the lower the jaw). The results are shown this way to adhere to the hypothesis that syllables with increased jaw displacement have greater prominence. Thus, the y-axis is labeled "Syllable Magnitude (mm)", since, according to Fujimura's C/D model hypothesis (1992), syllable magnitude is correlated with the amount of jaw displacement required to articulate each syllable. In the figures below, each syllable has a different magnitude, even though the vowel is phonologically the same. In our figures, ordinate scaling is set individually for each speaker, to better view the patterns of jaw opening.

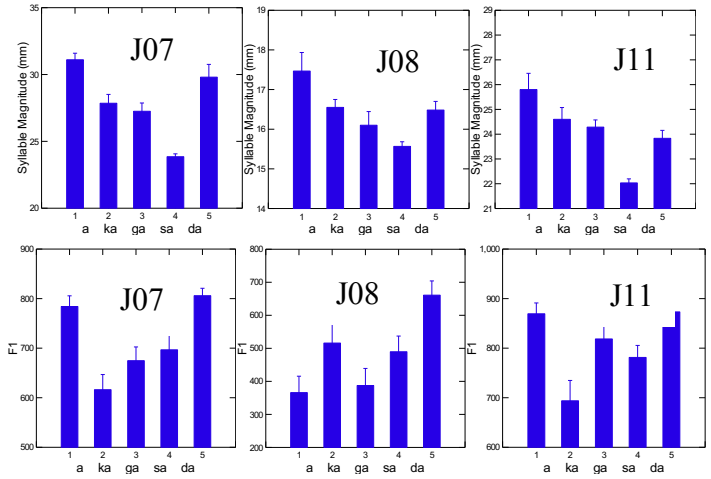


Figure 2: Syllable magnitude and F1 patterns for the Japanese utterance *akagasa da* for L1 speakers J07, J08, J11. The top row shows syllable magnitude patterns; the bottom shows F1 patterns.

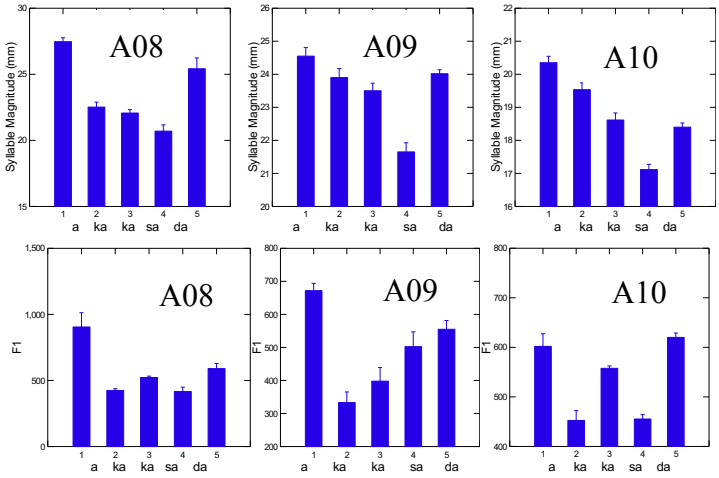


Figure 3: Syllable magnitude and F1 patterns for the Japanese utterance *akagasa da* for L2 speakers A08, A09, A10. The top row shows syllable magnitude patterns; the bottom shows F1 patterns.

Looking at the syllable magnitude patterns in Figure 2, we observe consistent patterns: the initial and final syllables of the utterance show comparatively large jaw displacement patterns, as

was reported in Kawahara et al. (2014). The correlations between jaw displacement patterns and F1 seem complicated; for J07, they seem to match except for the second and fourth syllables; for J08 and J11, there do not seem to be systematic correlations.

For this sentence type, the syllable magnitude patterns for the L2 speakers (in Figure 3) are strikingly similar to those of the L1 speakers from Figure 2, again showing utterance initial and final increased magnitudes. Interestingly, the F1 patterns seem to correlate with jaw displacement magnitudes for A08 (r^2=0.60) but not for the other L2 speakers.

3.2. L1 and L2 patterns of jaw displacement and F1 for *aka pajama da*

Looking at the syllable magnitude patterns for the L1 speakers in Figure 4, we again observe large jaw displacement for initial and final syllables. Like the *akagasada* sentence, no strong correlations were found between jaw displacement and F1 for the L1 speakers. Note also that jaw displacement is larger on the first syllable of *pajama* than on the second syllable. It could be argued that this is because in Japanese, *pajama* has a pitch accent on the first syllable; however, Kawahara et al. (2014) show that pitch accent does not affect jaw movement patterns.

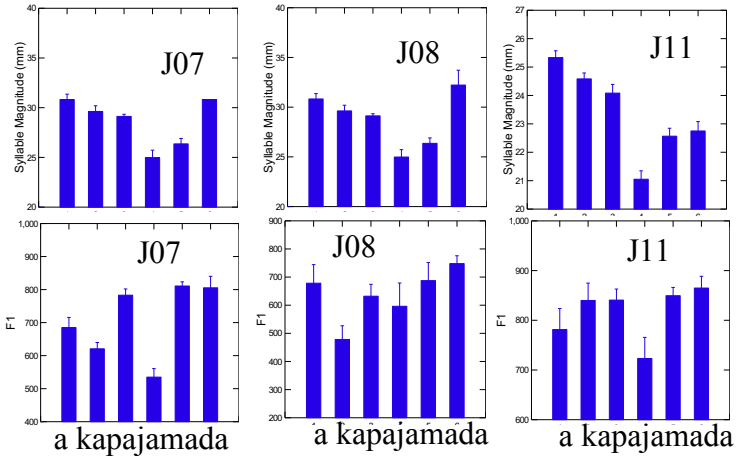


Figure 4: Syllable magnitude and F1 patterns for the Japanese utterance *aka pajama da* for L1 speakers J07, J08, J11. The top row shows syllable magnitude patterns, the bottom shows F1 patterns.

These utterances contain the English cognate word, “pajama”, and for the L2 speakers (in Figure 5), we see a different pattern of syllable magnitude than that for the L1 speakers, and different from the *akagasada* sentence as well. Specifically, [ja] has a larger jaw opening than the preceding [pa]—this is presumably due to negative L1 transfer (i.e., interference) from the lexical stress pattern of the English word *pajama*, where stress is on the second syllable. For cognates, such negative L1 phonetic transfer has been shown for VOT in Amengual (2012). It has also been shown

for stress assignment in Ghazali & Bouchhioua (2003), but negatively interfering from Tunisian participants' L2 (French) to their L3 (English) for French-English cognates.

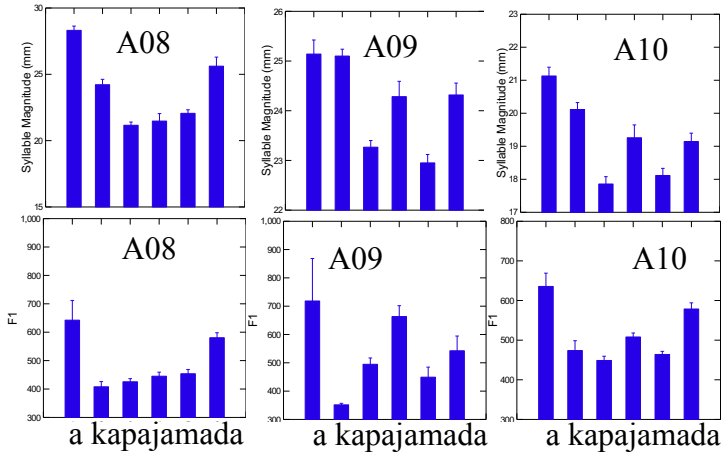


Figure 5: Syllable magnitude and F1 patterns for the utterance *aka pajama da* for English L2 speakers, A08, A09, A10. The top row shows syllable magnitude patterns, the bottom shows F1 patterns.

In Figure 5, the F1 patterns match the syllable magnitude patterns, with increased F1 on the middle syllable. One L2 speaker, A08, however, seems to show almost equal amounts of syllable magnitude as well as F1 values on the three syllables in *pajama*, thus endeavoring to avoid the L1 pattern of increased strength on the middle syllable. Indeed, for the participants in Ghazali & Bouchhioua (2003), the number of misplaced stresses was inversely proportional to their proficiency level, so it is possible that jaw movement here could be an indicator of L2 proficiency level. Regression analyses show positive relationships between jaw displacement and F1 for A08 and A10 (r^2=0.44, r^2=0.39, respectively).

4. Conclusion

The results from this small study suggest that English speakers fluent in speaking Japanese are relatively successful in articulating Japanese edge prominence characteristics of phrasal stress. Specifically, they produce increased jaw displacement patterns similar to Japanese L1 speakers at phrasal edges (both initial and final), thus indicating a lack of interference of English articulatory prosody on that of Japanese. However, in utterances which involve English loan words, there appears to be some negative L1 transfer. Specifically, the L2 speakers of Japanese, especially A09 and A10, show clear increases of jaw displacement on the English lexically stressed middle syllable of *pajama.* This contrasts with the Japanese L1 speakers who show larger jaw displacement on the first syllable of *pajama* than on the second syllable. It is interesting to note that the L2 speakers for the utterance with the English cognate also tend to show a strong relation between jaw displacement

and F1. The relation between jaw and F1 is a topic of continued exploration (see, e.g., Huang and Erickson 2019 for discussion of jaw and tongue interaction for prosody articulation).

Acknowledgements

Thanks to Jianwu Dang and Atsuo Suemitsu for use of and technical help with articulographs, and to the participants in the experiment. This work was partially supported by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C) #25370444 to the second author.

References

Amengual, M. (2012) "Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism", *Bilingualism: Language and Cognition* 15(3), 517–530.

Erickson, D. (2002) "Articulation of extreme formant patterns for emphasized vowels", *Phonetica* 59, 134–149.

Erickson, D., Suemitsu, A., Shibuya, Y., and Tiede, M. (2012) "Metrical structure and production of English rhythm", *Phonetica* 69, 180–190.

Erickson, D., Kawahara, S., Shibuya, Y., Suemitsu, A. and Tiede, M. (2014) "Comparison of jaw displacement patterns of Japanese and American speakers of English: A preliminary report", *Journal of the Phonetic Society of Japan* 18, 88–94.

Erickson, D., Kim, J., Kawahara, S., Wilson, I., Menezes, C., Suemitsu, A., and Moore, J. (2015) "Bridging articulation and perception: The C/D model and contrastive emphasis", *Proc. of 18th ICPhS*, Glasgow.

Erickson, D., Iwata, R., and Suemitsu, A. (2016) "Jaw displacement and phrasal stress in Mandarin Chinese", TAL 2016.

Fujimura, O. (1992) "Phonology and phonetics — A syllable-based model of articulatory organization" *Journal of the Acoustical Society of Japan (E)* 13, 39–84.

Ghazali, S. and Bouchhioua, N. (2003) "The learning of English prosodic structures by speakers of Tunisian Arabic: Word stress and weak forms", *Proc. of 15th ICPhS*, Barcelona, 961–964.

Hirata, Y., Whitehurst, E., and Cullings, E. (2007) "Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates", *Journal of the Acoustical Society of America* 121(6), 3837–45.

Huang, T. and Erickson D. (2019) "Articulation of English 'prominence' by L1 (English) and L2 (French) speakers", *Proc. of 19th ICPhS*, Melbourne.

Kawahara, S., Erickson, D., Moore, J., Suemitsu, A., and Shibuya, Y. (2014) "Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress", *Journal of the Phonetic Society of Japan* 18.2, 77–87.

Kawahara, S., Erickson, D., and Suemitsu, A. (2015) Edge prominence and declination in Japanese jaw displacement patterns: A view from the C/D model. Special issue on the C/D model, *Journal of the Phonetic Society of Japan* 19.2, 33–43.

Menezes, C. and Erickson, D. (2013). "Intrinsic variations in jaw deviations in English vowels", *POMA* 19, 060253.

Shport, I. (2016) "Training English listeners to identify pitch-accent patterns in Tokyo Japanese", *Studies in Second Language Acquisition* 38(4), 739–769. doi:10.1017/S027226311500039X

Smith, C., Erickson, D., and Savariaux, C. (forthcoming) "Articulatory and acoustic correlates of prominence in French: Comparing L1 and L2 speakers", *Journal of Phonetics* (special issue).

A4

複数人の英単語同時発声における音声の物理的評価と心理的評価

高野 佐代子（金沢工業大学　メディア情報学科）
tsayoko@kanazawa-it.ac.jp

1. はじめに

現在の英語発音における教育方法として、伝統的には教師が英単語を発音したものを複数人の生徒が繰り返す同時発声が行われている。本研究ではこの複数人の英単語同時発声の効果を科学的に調査する。

これまでに山田らは科学的英語上達法(2005)、各種のステップを踏んだトレーニング(牧野、2013)、また近年は一般的にも発音学習においてシャドーイングによる英語の改善の効果も数多く指摘されている。英語学習法の提案は数多く存在する。我々は予備的な調査により、山田らの科学的上達法のうち、発音時の詳細な舌運動の表示の効果について各種の検討を行ってきた。舌運動の表示を伴う英単語の発音教示により、英語音声単独教示に比べて、英語母国語話者による英語発音の評価値を向上させることができる可能性を調べている。一方で、英語発音を評価した英語話者の内観として、正しく発音できているにも関わらず、自信がなさそうに発音することがとても残念だった、というコメントを得た。これは、英語発音時の英語の知識や正解・不正解の問題でなく、スワレスら (2001) が指摘する「恥ずかしさ」などが起因しているといえる。

そこで本研究では、各種指導法における興味などの動機付けなどの個人差が関与する要因ではなく、一般に教室内で用いられれる複数人同時発声の効果を検証する。これは音声知覚・生成において知られている「ロンバート効果（Lane, 1971) 」を利用した音声の変化を英語学習の場面において利用することを想定する。「ロンバート効果」とは人は無意識に周囲の環境の聴覚的な音刺激に打ち勝つように音声を発することである。

実験 1 では複数人同時発声の音声の収集と分析、実験 2 では複数人同時発声時の音声の心理的評価を行う。これにより、単独発声と複数人同時発声の音声の違いについて比較する。

2. 実験 1 複数名同時発声の物理的評価

2.1. はじめに

本実験では、複数名同時発声時における発声者の音響パラメータを調査する。これに先立って、見本音声と複数名同時発声に使用する音声の両サンプルを準備する (2.2.1)。また、これらの音声を用いて同時発声実験を行い(2.2.2)、録音された音声を分析し音響物理パラメータを明らかにする(2.2.3)。

2.2. 方法

2.2.1 英語の音声サンプルの準備

使用する英単語は「日本人が最も苦手意識を持つ英語発音記号」(Ryne Richards, 2002)を参考に、日本人が最も苦手とする発音記号である音声と対立する/r/と/l/, /v/と/b/, /s/と/th/とした。選定した発音記号に対し、英単語のリストから「日本の高校生が学ぶ」かつ「比較的幅広く認知されている」英単語を抽出する。/b/として best, big, boat, /v/として vest, video, vote, /l/として lock, light, load, /r/として road, right, road, /s/として sank, sing, sink, /th/として thank, thing, think とした。

見本および自身の発声と同時に再生される音声は、言語データベースシステム FORVO から得た男性 5 名、女性 5 名計 10 名の英語音声サンプルである。英語の音声サンプルとして選ばれた基準は「母国語が英語である」、「公用語が英語の地域に在住している」、「英語教育に変わっている」という条件に合致するユーザがアップロードしたものから選択した。

2.2.2 同時発声実験および録音環境

被験者（発声者男性計 5 名）が無響室内においてヘッドフォン（MDR-CD900ST, SONY）を装着した状態で、見本の音声に続いて各種条件下で発声する。マイク（NT-2A, RODE）から口唇の距離は 50 cm とし、USB オーディオインターフェース（OCTA-CAPUTURE, Roland）を使用して被験者が発声した音声を録音した。

見本の音声に続いて、被験者は同じ単語の複数名同時発声を 3 回行う。上記にも示したように、使用する英単語は/b/-/v/、/l/-/r/、/s/-/th/を含む対立音声各 3 単語計 18 単語である。単独、2 名、4 名の条件はブロックでランダム順とした。複数名同時発声の音声は、男女同人数であり、男性 5 名女性 5 名中からランダムに選ばれる。これらの実験環境は数値計算統合環境 MATLAB のアプリケーション作成 GUI コマンドの guide を使用して作成した。

また、実験の終了にともなって、英単語の意味に自信があるか、英単語の発音に自信があるか、被験者 1 人で発音する環境に対し緊張したか、被験者 1 人で発音する環境に対し発音しやすかったか、複数人の音声と同時に発音する環境に対し緊張したか、複数人の音声と同時に発音する環境に対し発音しやすかったか、についてアンケートを収集した。

2.2.3 音声の分析

分析は数値計算統合環境 MATLAB の音響分析に特化した Auditory Toolbox を利用しラウドネス値を算出した。設定はフルスケールに対する相対ラウドネス単位（LUFS）で、Momentary Loudness (400 ms)とした。

2.3. 結果

2.3.1 ラウドネスの変化値

図 1 にラウドネスの変化が一人に比べた変化を示す。ラウドネスの値の評価では、それぞれの項目で被験者 1 人が発音する環境と複数人の音声と同時に発音する環境値の差を出し、被験者ごとの差の平均値と比較する。平均値より差が大きい場合は評価「＋＋」、平均

値より差は小さいがプラス値の場合は評価「＋」、差が平均値より小さくマイナス値の場合は評価「－」とした。

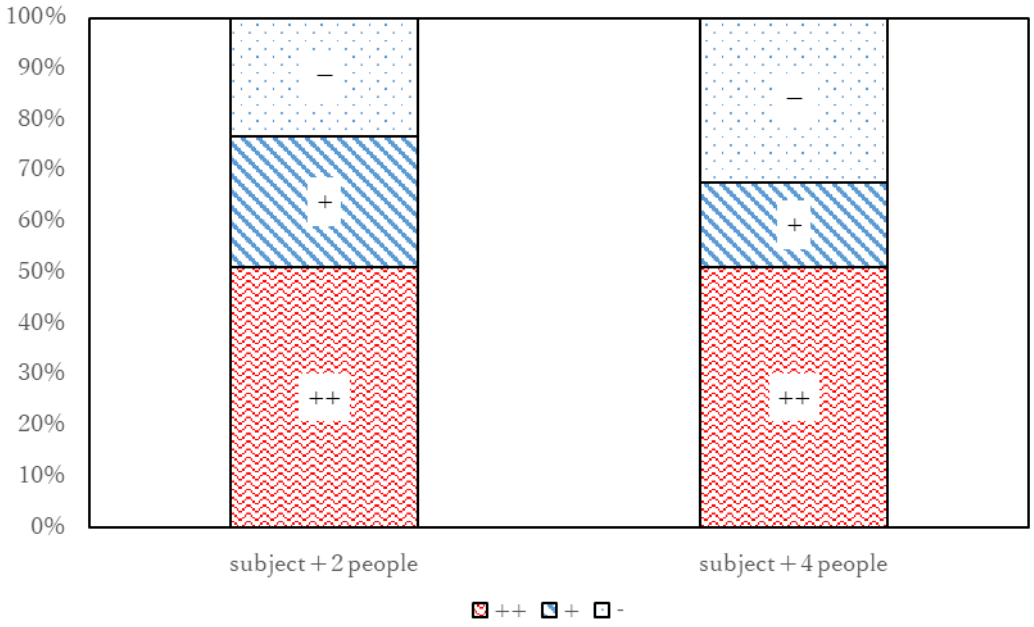


図 1 単独発声と比較したラウドネス値の変化

一般に、ヒトは発話時に環境音に打ち勝つように発声することが指摘されており、このような現象は「ロンバート効果」と命名されており、ここでも同様の結果が見られた。1 人で発音する環境の音声よりラウドネスの値にあきらかに大きくなった音声は 2 名の場合、4 名の場合で双方とも 51%、若干おおきくなった音声は 2 名の場合で 26%、4 名の場合で 17%であった。

総じて、単独に比べて複数名の同時発声の環境の方が約 70%の発音にいてラウドネスが大きかった。

2.3.2 アンケートによる発声者の内観報告

まず英単語の意味、および発音に対して自信があるかについて尋ねた設問に関して、1 人のみが自信があり、他は自信がない、やや自信が無いとの回答であった。意味に関しては自信があっても、発音には自信がないという人も見られた。

また環境の違いによる設問に関しては、単独の場合には緊張する、発音しにくい、などの回答があったが、複数の場合には緊張の度合いがやや弱まり、発音しやすい傾向が見られた。ただし、タイミングなどを気にすることによる発音のしにくさも見られた。

2.4. 考察

約 70%のサンプルにおいて、複数名同時発声によりラウドネスが上昇する効果が見られた。変化の小さい発声者は単語の意味、発音ともに「自信がある」と答えた被験者であっ

た。英語発音における同時発声によるラウドネスの変化への効果は、ロンバート効果と同様に多くの人に見られるが、個人の自信の有無にも関係している可能性がある。

3. 実験 2 複数名同時発声音声の心理的評価

3.1. はじめに

実験 1 の発声実験では、被験者 1 人で発音する場合や複数名で同時発声するときの音声を収録し、loudness の変化を比較した。一般に、loudness に影響を受ける被験者が多く、複数名で発音することにより大きな音声になるものが多く見られたが、loudness に影響を受けにくい被験者も見られた。ここでは、複数名発声音声は英語発声において聴覚心理的に効果があるか、確認する必要がある。ただし環境による発音の違いは個人内で類似しており、話者の個人差の方が大きい。そこで一対比較を用いて、被験者内において 1 人で発音、2 名の音声と同時に発音、4 名の音声と同時に発音に対する評価を調査する。

3.2. 方法

3.2.1 シェッフェの一対比較法: 中屋の変法

一対比較の方法としてシェッフェの一対比較法: 中屋の変法を行い、3 つの環境の音声に対する印象評価の比較を行う。中屋の変法は「順序効果の考慮不要」、「各被験者が全部の対比較評価をする」という条件から選択した。具体的には実験 1 の被験者 5 名のうち大きく影響が見られた 2 名と影響があまり見られなかった 1 名の/l, r/の/light, right/と/s, th/の/sank, thank/の各環境音声データを使用する。

比較項目は、①どちらの方がより自信があるか、②どちらの方がより明確であるか、③どちらの方がより音声として高いか、④どちらの方がより受ける印象として明るいかとし、5 段階の両極尺度で評価させた。

3.2.2 実験環境

日本人成人男性の被験者 10 名が無響室内においてヘッドフォン（MDR-CD900ST, SONY）を用いて、上記に述べた英語音声の聴覚印象を評価する。英語発音の優劣の評価は行わない。

3.2.3 統計手法

被験者一名で発音した音声を A、2 名の音声と同時に発音した音声 B を、4 名の音声と発音した音声を C としてスコア集計、心理尺度値のプロット等を行う。それぞれの評価対象のスコア集計には R 言語と R Studio に ScheffePairedComparison.R（Shigenobu Aoki）をダウンロードし使用する。心理尺度値、yardstick 値の算出は、評価対象数は 3、評価項目数は 4、5 段得点付けとした。心理尺度値のプロットには数値計算統合環境 MATLAB を使用し、複数名の発声環境によって「自信がある」や「明確である」の評価の違いを可視化する。

3.3. 結果

実験 1 の発声実験のうち 1 名分について、一対比較データの心理尺度のプロットを図 2 に示す。

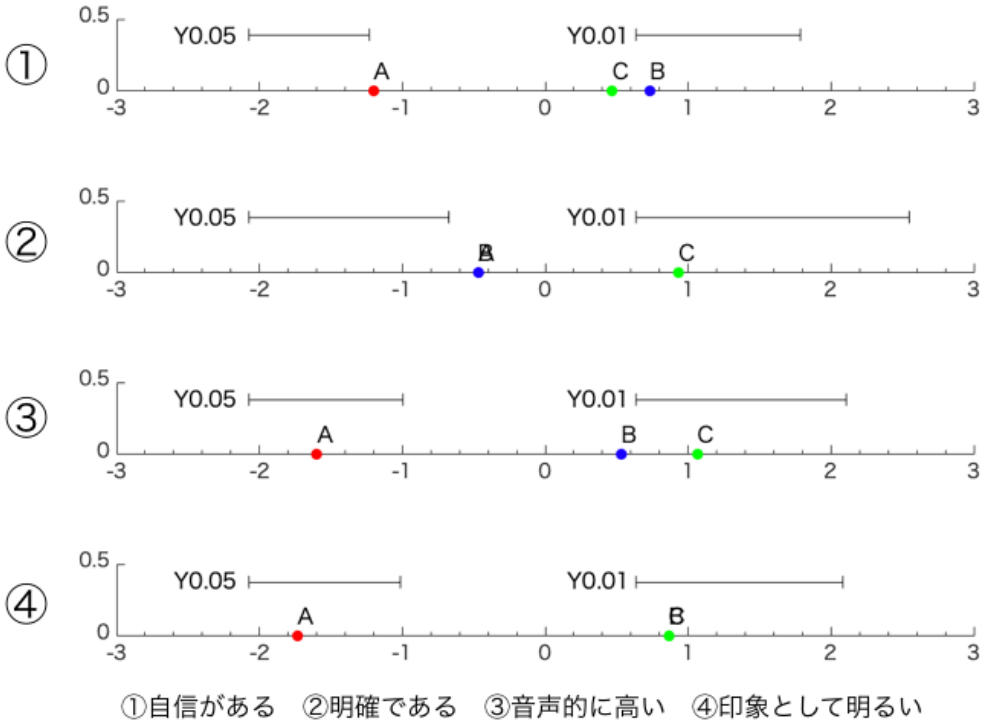


図 2 発声者 1 名の音声に対する一対比較の心理尺度プロット

全体的に被験者 1 人で発音する環境(A)より複数人の音声と同時に発音する環境(B, C)で録音した音声の方が「自信がある」「高い」「印象として明るい」と評価されている。一方で「明確である」に関しては同程度であった。この傾向は他の 2 名の発声者でもほぼ同等であった。特に実験 1 でラウドネスが同程度であった発声者においても、程度は小さいものの、同じ傾向が見られた。

3.4. 考察

複数名による発声環境による音声の違いを一対比較による聴取実験により調べた結果、自信がある、高い、印象として明るい、などの結果を得た。一般的なロンバート効果などによるものに加えて、英語発声という特殊な環境における効果を検討した。

「日本人学習者の英語発音に対する学習態度について」（アーマンドら，2001）で挙げら

れている日本人学生が持つ正しい発音を行うことへの「恥ずかしさ」や「冷やかされる」といった心理的要因を排除する一つの手法として、今回行なった複数人との同時に発音するということが効果的であると考えられる。

一方で、発音の明確さの差は小さかった。同時発声によって、英語の発音自体が良くなるという効果は小さい可能性があり、教室内における指導についてはさらに工夫が必要である。これに関しては、さらに発声者に対する発音等の指導が必要であると考えられる。

4. 結論

通常教室内で行われる英単語の複数名同時発声による効果について、物理的評価と心理的評価の両側面から行った。その結果、環境に打ち勝つために声が大きくなるロンバート効果と類似した効果が約 70%の音声に見られた。心理的評価により、特に英語発声という状況では、複数名同時発声により、自信がある、高さ、明るさなどの項目への効果が見られた。

謝辞

本研究は、2018 年度金沢工業大学メディア情報学科卒野村洸太君が中心となって研究を行ったものである。また新設なアドバイスを多くいただいた ATR Learning Technology 山田玲子氏、上智大学北原真冬氏、法政大学田嶋圭一田氏に心よりお礼申し上げる。

参考文献

山田恒夫, 英語スピーキングの科学的上達法, 講談社, (2005).

スワレス アーマンド、 田中 ゆき子, “日本人学習者の英語発音に対する学習態度について”, 新潟青陵大学紀要第 1 号,(2001).

牧野眞貴,“学生が効果的に感じる英語発音トレーニングの実践報告”, 外国語教育フォーラム第 12 号, (2013).

Lane, H. and Tranel, B, “The Lombard Sign and the Role of Hearing in Speech”, J. Speech Hear. Res., Vol. 14, pp. 677–709 (1971).

Ryne Richards, “日本人が最も苦手意識を持つ英語発音記号”, 九州女子大学紀要, (2002).

A5

英語音声教育における TTS 合成音声の活用とその問題点

東　淳一（神戸学院大学）
azuma@gc.kobegakuin.ac.jp

1. はじめに

今日私たちの生活の周りにはTTS合成音があふれている．例えば速報性が要求される高速道路のハイウェイラジオも最近TTS合成音になったし，合成音により読み上げられた音をそのまま放送している日本のFM放送局も存在する．新幹線のアナウンスにもTTS合成音が一部利用されている．今日の社会において，能率的に音声情報を発信するためにかなり多くの部分でTTS合成音が利用されてきており，私たち自身これがTTS合成音であると気づいていないサービスも多い．今後もTTS合成音の利用はますます拡大すると考えられる．

実は音声言語が必要とされる場面は，教育分野においても非常に多い．もちろん我々の日々の講義も音声を使って行っており，また英語教育においてはターゲット言語の音声そのものが大変重要な教材となり使用されている．もちろん教員本人も授業中に英語を発するが，ある程度自動的かつ大量に文型練習を口頭で行わせる，音声の聞き取り練習をさせる，スピーチやプレゼンのモデル音声として音声を提示するなど，生身の教員だけでは対応できないケースも多い．また，ヒアリングの試験では英語のネイティブスピーカーによる音声が必要となり，その場合，アメリカ英語のみならず世界各地の英語の音声が必要とされることもある．現実に，TOEICのリスニングセクションでは，アメリカ英語，イギリス英語，カナダ英語，そしてオーストラリア英語が利用され，受験生は異なるタイプの英語音声に対応できなければならない．大学院生や研究者については，国際学会において英語で発表するケースも頻繁にあり，その場合にはモデルとなる英語音声を聞きつつ何度も発話練習するという必要に迫られることが多い．ビジネスマンの場合にも，英語でのビジネスプレゼンの練習，あるいは英語による交渉のための模擬練習も欠かすことができない．やはりプレゼンのモデルとなる英語音声や，交渉のさまざまな場面を想定した英語での想定問答がネイティブスピーカー発話による音声として提供されれば，非常に有益である．発話練習，スピーチの練習が必要であっても，まずはモデルとなる適切な英語音声のインプットが存在することが重要である．さらに学習者の能力と聴衆の属性，話される内容のジャンルなどに対応した適切な韻律の英語音声が学習者にモデル音声として提供されることが必要である．ネイティブスピーカーによる発話の録音にかかるさまざまなコストを考えれば，このような場面においてもTTS合成音声の活躍が大いに期待される．

上記のようなニーズを考慮に入れ，本研究では，特にAmazon Pollyに代表される，クラウドベースのTTSサービスの実態とその利用方法について述べ，さらに学習者のレベルや産出される英語音声のジャンルなどに対応する適切な合成音声を作り出すには，TTS合成音作成過程で，どのように韻律を制御しなければならないのかについても解説する．

2. ネイティブスピーカーにモデル音声の録音を依頼する際のコストと問題点

2.1. 人材確保の難しさと金銭的コストの問題

英語教員がネイティブスピーカに録音を依頼し教材を作る，あるいは英語のリスニングテストを作るなどの作業を行いたい場合，いくつかよく問題になる点がある．その一つはコストの問題である．ネイティブスピーカーの先生に録音を依頼するということになるとかなりの録音のための時間を要することになり，またそのためにある程度の謝金を用意するということになるが，その額についてもそう安くはできない．また，これは地域にもよるが録音に対応可能なネイティブスピーカーを確保することは容易ではない．

2.2. 技術的問題

ネイティブスピーカーに録音を依頼する場合，外部からの雑音が遮断された録音室が必要となるが，そのような環境が教育施設に整備されている可能性はかなり低い．また高い謝金を払ってネイティブスピーカーに録音をしてもらったものの，録音終了後にほんのわずかな録音漏れがあった，あるいは録音のミスがあったことがわかった場合，いかにそれを修正するのかも大きな問題となる．再録が可能になったとしても，録音機材の録音レベルや録音室内での話者が座る位置の違い，あるいはマイクの位置の違いや話者の健康状況などによっても録音品質や雑音のレベルが大きく変わってくる．このため，後に録音した音声をサウンド編集ソフトを使って最初に収録した音声データに貼り付けても，不自然な音声に仕上がってしまう可能性がある．

2.3. ネイティブスピーカーの問題

ネイティブスピーカーについても，依頼者の思うように素材を発話してくれるとは限らない．当然収録前の打ち合わせにかかる時間も多くなり，また修正を依頼すると，ネイティブスピーカー自身が依頼者の考える韻律は不自然であって自分が正しいと主張する場合もあり，なかなか想定しているような理想の音声の収録ができないこともある．

3. TTS 合成音声を使用する利点と利用シーン

3.1. TTS 合成音声使用の利点

テキスト入力を行い，ポンとクリックすれば音声が合成されるという利便性だけではなく，TTS 合成音声を利用することにはさらに複数の利点がある．もちろんもっとも重要な点は音声教材の作成が極めて簡単になるということである． 2 番目として，音声はデジタルデータとして電子的に蓄えられるため，その音声素材の共有は極めて容易になることがある．つまり他の教員，学習者との音声ファイルの共有がきわめて容易になり，またその配布も簡単になる． 3番目にあげるべき点は音声データ編集の容易さである．かつてのアナログ音声データを編集あるいは修正しようとすると大変な労力と時間がかかった．しかしデジタルデータであれば適切なソフトにデータを瞬時に読み込み，編集・修正が終われば瞬時にデータを書き込むことができる．また4つ目にあげるべき点として，かつてのオーディオテープのようにメディア素材そのものがさほど短時間に劣化することがないとい

うことがある．このため保存したメディアの経年変化についてはさほど神経質になる必要がない．

3.2. 英語教育での TTS 合成音声の利用シーン

TTS 合成音を英語教育の現場で活用する場合，想定されるシーンとしては以下のようなものが考えられる．

- リスニング練習用のモデル音声，指示のための音声として．
- リスニングテスト用のモデル音声，指示のための音声として．
- シャドーイングやオーバーラッピング練習のためのモデル音声として．
- パブリックスピーチやプレゼンテーションの練習のための音声モデルとして．
- 対話練習用の音声素材として．学生の発話部分を無音にしておき相手方の音声のみを作成しておく等．
- 動画制作において．ナレーションやアフレコ，あるいは TTS の複数の音声による擬似的な対話音声の作成と挿入等．
- 教員の研修用の音声素材として．特に小学校での英語教育のための教員研修等．

リスニング練習，あるいはリスニングテストについては Moodle などの LMS を使い，オンラインで実現することが可能である．次の図 1 は，TOEIC のリスニング問題で絵を見てその描写として正しい英語の発話を選ぶ問題を，サンプルとして Moodle で作成してみたものスクリーンショットである．

図 1：Moodle 上でのリスニング問題のスクリーンショット

図中のプレーヤを再生すると１～４の４種類の音声が流れ，写真の描写として正しいものを選ぶ形のものである．なお，新しい TTS 合成音活用シーンとしては，動画でのナレーションやアフレコの用途が考えられる．音声の収録以上にビデオの収録は時間がかかり，さらに何か不備が後で見つかったとしても撮り直しが大変困難である．そういうケースに備えて，話者の姿を写さなくても良いシーンや，静止画を次々と表示しバックでナレーションを入れるだけで済むような場面も作っておけば，音声については TTS 合成音で十分に間に合う．また，登場人物が日本語で話している場面で英語での音声の吹き替えを行う必要がある場合にも，TTS 合成音の使用が便利である．

4. Amazon Polly の利用について

ほんの２，３年ほど前までは TTS といえばスタンドアロンタイプのソフトが普通であり，ソフトには TTS 音声は含まれておらず，同時に好みの TTS 音声を購入しなければならないという場合も多かった．東（2018）で報告したように，現在では Amazon，Google，Microsoft がそれぞれクラウド上で Web サービスを展開しており，それぞれにおいて TTS のサービスも利用できる．この３社のうち，Amazon Polly が Web 上の GUI ベースのインターフェース上で音声合成を実現できるため，特に気軽に使える．紙面の関係で手続きは省略するが，使用に当たってはあらかじめアマゾンウェブサービス（以下 AWS）に登録をしておく必要がある．Amazon Polly をキーワードに検索し，登録されたい．なお，ログイン後には AWS マネジメントコンソールの画面に移動するので，そこで「サービスを検索する」の部分から Amazon Polly もしくは単に Polly で検索をかける．すると，以下の図２のようなコンソール画面に到達する．すでにテキストを入力してあるが，中央の四角い部分にテキストを打ち込み，「音声を聞く」スイッチをクリックすれば音声を再生し，「ダウンロード MP３」をクリックすると MP3 形式で音声ファイルをダウンロードできる．2019 年７月末からは，24000Hz サンプリングの高品質音声が提供されており，その場合にはエンジンをスタンダードではなくニューラルに切り替える．Amazon Polly の課金情報は，https://aws.amazon.com/jp/polly/pricing/ に記載されているが，原則１件あたり 1,000 文字のリクエストを 1,000 件行った場合（つまりテキストの長さにして合計 100 万文字）スタンダード音声が４米ドル，ニューラル音声が 16 米ドルであり，大変安価である．合成音について使用目的の制限はない．

通常，合成音はニュース音声をモデルとして合成される．このため発話速度はかなり早く，韻律についてもニュース番組のアナウンサーのようなスタイルになり，イントネーションについてもあまり大きな F0 の起伏はない．フレーズの長さも長く，ピリオドでのポーズの長さもさほど長くはない．このままでは日本の英語教育用の英語音声としては利用が困難かもしれない．しかしながら，Amazon Polly では合成音作成者が合成音にお好みの韻律を付与できる．この場合には，コンソールの「プレーンテキスト」ではなく「SSML」のタブを選択し，そちらにテキストを入力する．

図2：Amazon Polly のコンソールのスクリーンショット

下の図３が SSML 対応のコンソールに必要な SSML タグを打ち込んだ場合のスクリーンショットである．

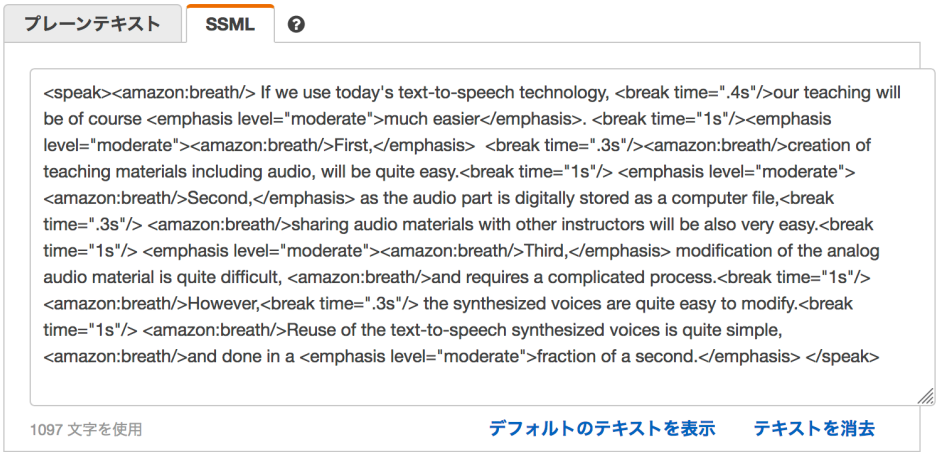


図3：Amazon Polly の SSML モードでのコンソールのスクリーンショット

SSML タブ右横のクエスチョンマークをクリックすると，SSML に関する説明が表示される．また， https://docs.aws.amazon.com/ja_jp/polly/latest/dg/supported-ssml.html では Amazon Polly が現段階でサポートする SSML タグがその使用法とともに示されている．実際のとこ

ろ，SSML で韻律のファインチューニングを行うことで，それぞれの教育シーン，あるいは学習者のニーズにうまくマッチした韻律をもつ合成音を作成することが可能である．

なお，Azuma(2008)あるいは Azuma(2010)で述べたように，従来のスタンドアローン形式の TTS 合成ソフトの利用については，合成された音声ファイルの利用についての制約が多く，作成したファイルを学習者に渡したり，誰もがアクセスできる Web サイトで音声のストリーミングを行ったりということは，ライセンスの関係で大変困難であった．この点，最近登場したクラウドベースの TTS 合成サービスは僅かな課金はあるものの，作成された音声ファイルの配布についての制約がなく，英語教員にとっては大変使いやすい環境が整ったといえるだろう．

5. おわりに

クラウドベースの TTS 合成音作成サービスの Amazon Polly を中心に，TTS 合成音の英語教育シーンでの活用法について述べた．たしかに人的コスト，時間的・金銭的コストを大幅に削減しつつネイティブスピーカーにかなり近い英語音声を作成できる，そしてわずかな課金があるものの作成された音声ファイルを自由に配布，活用できるという点で，これらクラウドベースの TTS サービスは私たちにとって福音である．しかしながら，SSML タグを巧みに打ち込み，試行錯誤を繰り返しつつ，それぞれの教員シーンや学習者のニーズにマッチした韻律をもつ合成音を作成することは，思ったほど容易ではない．

今後の英語教育での TTS 合成音声の利用研究の重点は，同じサービスの同じボイスを使ったとして，いかに教員のニーズにマッチした韻律をもつ音声をそのボイスから作り出すか，つまりいかに SSML タグを駆使するのか，そしてさらにその次には， いかに SSML を駆使して教育シーンにマッチした TTS 合成音を作成できる「TTS 合成音プロデューサー」を養成することができるかといった点に移行するであろう．

参考文献

Azuma, J. (2008) “Applying TTS technology to foreign language teaching,” *Handbook of Research on Computer-Enhanced Language Acquisition and Learning*, Zhang, F. and Barber, B. eds, IGI Global, Hershey, 497-506.

Azuma, J. (2010). “A quantitative evaluation of the quality of Text-To-Speech (TTS) engines: toward the application of TTS technology to TEFL,” *Language Education and Technology (LET 50) Conference Proceedings*, Yokohama, 232-233.

東淳一 (2018) 最新 TTS 合成音声の外国語教育現場での活用．2018 年度外国語教育メディア学会全国研究大会．2018.8.9.

A6

日本語における発話リズムの異常性について
—運動障害性構音障害の発話をとおして—

難波 文恵（川崎医療福祉大学大学院）
fnamba@hotmail.com

1. はじめに

1.1. 発話リズムの異常

「リズム」という用語は発話の障害においても用いられる。運動障害性構音障害[1]や吃音では「リズムの異常」があるとされる(廣瀬 2001)が、筆者は言語聴覚士としてその評価の曖昧さを実感する。例えば福迫ら(1983)の麻痺性(運動障害性)構音障害の評価票は「バラバラという印象」「不規則にくずれる」「不自然にとぎれる」、また小澤ほか(2013)の吃音の非流暢性分類も「不自然に引き伸ばされる」「話者の発話の流れにおいて不自然」と主観的に表現される。ここに何らかの客観的指標が示されれば、より具体的に発話症状を記述し、他の症例との比較や経過を追うことができ、より具体的な訓練の方向性が設定できるだろう。

1.2. リズムの定義

リズムの対象となる領域は広範囲でバリエーションに富む。フレス(1987)は「リズムを研究することは困難な課題である。正確でかつ一般的に受け入れられるようなリズムの定義というものが存在しないから」とした上で、「リズムは継起する事象の秩序だった特性であり、この秩序は心で感じられ、知覚される。リズムは心的構築によって生まれるものである。次に何が起こるかを予期できる時に、リズムが存在する」としている。この定義に基づき、本研究ではリズムを、物理的現象に対して受け手の心的構築によってうまれる秩序の知覚として捉える。リズムが物理的に存在するのではなく、物理的事象に連続性や群といった秩序を認め、そこにリズムがあるとするのは、ヒトの認知能力だとする。

1.3. 言語リズムの二側面　　－「要素の規則正しい繰り返し」と「快さ」－

窪薗(1993)は、リズムは「よどみなく流れる」ことを意味し、そのためには「何か一定の構造が規則的に繰り返し起こらなくてはならない」とある。そしてリズムは「このような繰り返しであり、その繰り返しから生じる心地良さ」とする。また杉藤(1997)は、呼吸や脈拍や体の動きと同様に、「言葉も何らかの基準によるリズムを保ち、その場合に快く響く」とする。つまり、言語リズムには二つの側面がある。①要素の規則正しい繰り返しと、②そこから得られる快さである。①はそれを言語リズムの基本とした上で、「等時性」の議論を含めて、盛んに検討されている。しかし②の検討は十分でないように思われる。②の研究対象の多くは詩歌や童謡歌詞や現代詩であり、自然発話のリズムの快さではない。

[1] 運動障害性構音障害（dysarthria）とは「ことばの生成に関連した運動を制御する筋・神経系の異常に起因する構音の障害」（廣瀬 2001）のこと。かつて麻痺性構音障害とも呼ばれた。

1.4. リズムの快さ　　－「リズムの異常性」を通して見るもの－

日本語のリズムは「機関銃リズム」とも呼ばれる(金田一 1967)が、それは日本語母語話者ではない者による日本語リズムの印象である。本研究は、母語としてのリズムの印象を検討する。母語の「リズムの快さ」とは、「リズムの異常性」がない状態を想像的に意識するものではないか。「リズムの異常性」がない状態とは「流暢で自然な発話」といえるが、「流暢」「自然」も「非流暢」「不自然」があってはじめて意識されるだろう。よって「リズムの快さ」と「リズムの異常性」は相補的関係にある。氏平(2008)は「逸脱と思しきものから垣間見えるものを手がかりにして、背景にあるものを考察し、隠れている真相、すなわち正常とは何か、逸脱とは何かを究明する」とあるが、本研究のアプローチも同様である。

2. 目的

リズムの異常な発話と正常な発話の比較をとおして、聞き手にリズムの異常性を感じさせる時間的要因を探る。そして異常性を感じせない発話に必要な客観的指標を提示する。

3. 対象

「麻痺性構音障害の評価用基準テープ」(日本音声言語医学会)と「標準ディサースリア検査スピーチサンプル」(インテルナ出版)に収録されている、健常者 6 名と構音障害患者 6 名の音声データを用いた。対象者の情報を表 1 および表 2 に示す。

表 1: 健常者の情報

	年齢	性別
健常者①	22	女性
健常者②	25	男性
健常者③	46	女性
健常者④	45	男性
健常者⑤	66	女性
健常者⑥	69	男性

表 2: 患者の情報

	年齢	性別	疾患
患者①	54	男性	パーキンソン病
患者②	54	男性	パーキンソン病
患者③	66	男性	パーキンソン病
患者④	56	男性	脳梗塞
患者⑤	68	男性	脊髄小脳変性症
患者⑥	70	男性	脊髄小脳変性症

表 3: 患者の発話特徴

患者①	項目18「繰り返しがある」レベル2　　その他の特徴:「速すぎる」 発話速度が速い。音と語の繰り返しがある。(例「たびびとの」「ふきたてて」)子音が不明瞭。分節されていないモーラがある。
患者②	「声の高さの異常(高すぎる)」段階1　　ディサースリアのタイプ:運動低下性 発話速度が速い。その一方、ポーズが長すぎる。子音が不明瞭。分節されていないモーラがある。
患者③	項目6「声の翻転」レベル2　　その他の特徴:顕著な「速さの変動」「繰り返しがある」 語頭でつまっている。音と語の繰り返しがある。子音が不明瞭。分節されていないモーラがある。
患者④	「発話速度の変動」段階1　　ディサースリアのタイプ:UUMN 発話速度がやや遅い。「ぬがせた」が速すぎる。語尾が強く、高くなる。流れていかない。(例「こ／と／に」)
患者⑤	「発話速度の変動」段階2　　ディサースリアのタイプ:失調性 発話速度が遅い。一音一音がバラバラでつながらない。流れていかない。(例「たい／ようが」「しまし／た」「はじめ／ました」)
患者⑥	「発話速度の変動」段階3　　ディサースリアのタイプ:失調性 発話速度がやや遅い。「ました」が速すぎる。ポーズが多い。一音一音がつながらない。(例「か／ち／と」「はじめ／ました」)

患者の発話特徴は表 3 のとおりである。各患者ごと、サンプル記載の評価(上段)と、発表者の聴覚的印象評価(下段)を示す。全ての患者の発話で、リズムの異常性の印象を受けた。リズムの異常性の印象とは、福迫ら(1983)の評価票で、項目「速さが変動する」、「音・音節がバラバラに聞こえる」、「音・音節の持続時間が不規則に崩れる」、「不自然に発話がとぎれる」に該当する。発話速度は、患者①②は速く、③は変動し、④⑤⑥は遅く感じられた。

4. 方法

音読音声(「北風と太陽」冒頭 4 文)を、SUGI Speech Analyzer Version 1.07 で可視化し、日本語リズムの印象に時間的要因として関わると考えられる物理的現象①~③を分析した。まず①物理的モーラ長(音韻論的単位モーラの物理的関連量)の全体的特徴や隣接 2 モーラ間の特徴を捉えた。次に一段階上のレベルとして②発話区分ごとの平均モーラ長を調べた。更に無音部分に注目し③ポーズの頻度、位置、持続時間長を調べた。

5. 結果と考察

5.1. 物理的モーラ長

図 1 に示したように、健常者の物理的モーラ長は 50~355ms の間にあった(最小値 50~65ms、最大値 225~355ms、平均 119.3~151.9ms)。最小値は全て、無声摩擦子音から始まり母音の脱落したモーラだった。最大値は、文節末か、特に大きな切れ目の直前のモーラだった。一方、患者の物理的モーラ長は 35~533ms の間にあった(最小値 35~110ms、最大値 242~533ms、平均 107.7~301.4ms)。患者は、文節末だけでなく、文節中にも長いモーラが出現していた。

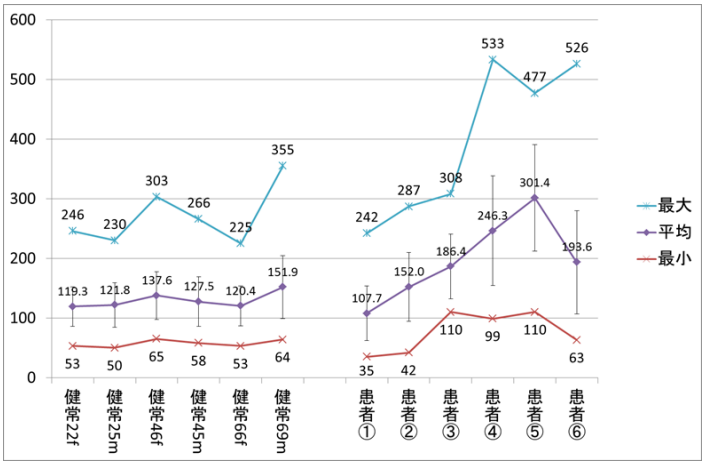


図 1: 物理的モーラ長の最大値、最小値、平均値 (ミリ秒)

ただし物理的モーラ長に影響を与える要因は様々あり(匂坂 1992)、環境の異なる各モーラの持続時間長を一括して検討するわけにはいかない。また要素の規則正しい繰り返しは、要素と要素の相対的関係において判断される。そこで最も近い位置にあるモーラ、つまり隣接 2 モーラの関係を調べた。

5.1.1. 隣接２モーラの物理的モーラ長の差

隣接２モーラの物理的モーラ長の差は、健常者では 0~235ms、患者は 0~314ms だった。各２モーラごとで、箱ひげ図の外れ値となる値を検出し、それらを「逸脱値」と判断した。

結果は図 2 のとおりである。隣接２モーラの物理的モーラ長の差が小さくても逸脱値となる場合[2]、大きくても逸脱値にならない場合[3]がある。隣接２モーラの差が「0」に近いほど等時性は高いわけだが、隣接２モーラの差の大小が、常に逸脱値か否かの判断と関連しているわけではない。とはいえ、隣接２モーラの物理的モーラ長の差が 235ms より大きい値は、そもそもそのような値は健常者の発話では出てこないが、患者の発話で出てくると必ず逸脱値となった。したがって、隣接２モーラの物理的モーラ長の差は、およそ 240ms 辺りを境界として、それ以上の差がある場合は必ず逸脱値と判断されるといえよう。

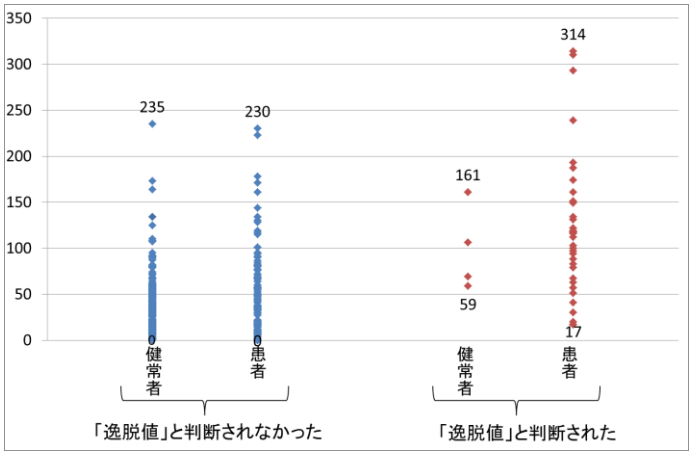


図 2:「逸脱値」と判断されなかった/された値 (ミリ秒)

5.1.2. 隣接２モーラの物理的モーラ長の比

隣接２モーラの物理的モーラ長の比は、健常者 1.00~2.96、患者 1.00~4.15 だった。比が 3.0 以上の隣接２モーラは、健常者はないが、患者は 2 割以上みられた。

5.2. 発話区分ごとの平均モーラ長

平均モーラ長は、健常者が最小 80~99ms、最大 147~220ms、患者が最小 67~239ms、最大 175~369ms で、発話区分ごと同様の長短パタンがみられるが、変動範囲は患者の方が大きい。健常者も患者も、平均モーラ長は発話区分内のモーラ数と負の相関がみられた。

5.3. ポーズ

ポーズ頻度(ポーズ回数／総モーラ数)は、健常者 8~10%、患者 15~23%で、患者は健常者よりも２倍近く多くポーズを挿入していた。ポーズ位置は、健常者は文と節の前後(文間、従属節と主節の間、引用文の前後)、主部と述部の間、文または節の初頭語の後ろのみだった。一方患者は、健常者と同じ位置の他に、文節内(形態素内、複合語内、動詞の複合語内、

[2]「まきつけました」の「ま→し」は、「ま」より「し」の方が短くなり、その差はマイナス値が自然である（「し」が無声化するため)。よって患者②の値「＋17」は逸脱値となった。

[3]「あるひ」の「る→ひ」で、健常者⑥の「235ms」は逸脱値とならない。「はじめました」の「め→ま」でも，患者⑤の「-230ms」は逸脱値とならない。

格助詞の前、連語内)や、IC の強い連結部分、文末の述語の前にもあった。

ポーズ長を図 3 に示す。健常者では、文間は 1200ms 以下、文内は 800ms 以下だった。患者では、文間でも文内でも 1800ms 以上があり、文内ポーズが文間ポーズよりも長い患者もいた。

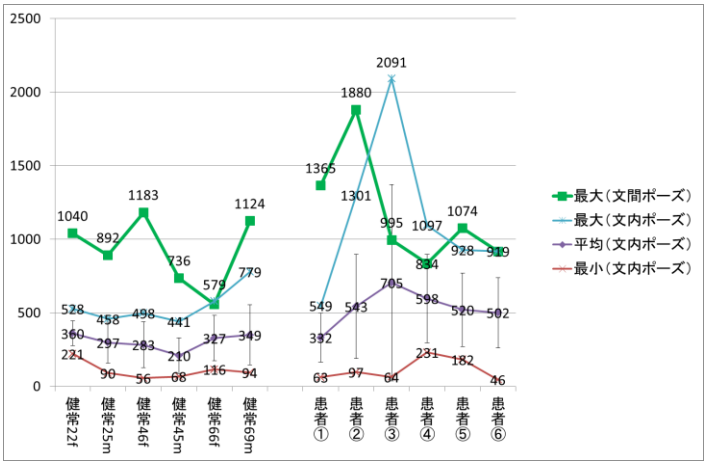


図 3: **文内ポーズ(最大値, 最小値, 平均値)と文間ポーズ(最大値) (ミリ秒)**

5.4. リズムの異常性の印象のメカニズム

フレス(1987)は「ある音の持続時間または音と音の間の時間間隔がめだって長くなると、そこで一つの群が終わりになる。この長くなった時間の長さによって、二つの隣り合うパターンの区別ができる」とし、「このような時間の延長により二つの群の間に切れ目が生じるが、われわれはこれを＜間＞[4]と呼ぶ」とする。そして＜間＞の長さは 1800ms 以上になると、群と群の繋がりを感じられなくなるとされる。

5.1.1 と 5.1.2 で述べたとおり、隣接２モーラの物理的モーラ長が約 240ms 以上の差、あるいは 3.0 以上の比となると、物理的モーラ長の急激な延長と判断される可能性が高い。つまりここで有音部分の＜間＞が生じる考えられる。＜間＞は統語的・談話的に合理的な位置に生じれば、群の切れ目を明示し、聞き手の理解にとって有効である。しかし＜間＞が統語的・談話的に不合理な位置に生じると、バラバラな感じ、すなわちリズムの異常性の印象をもたらすだろう。

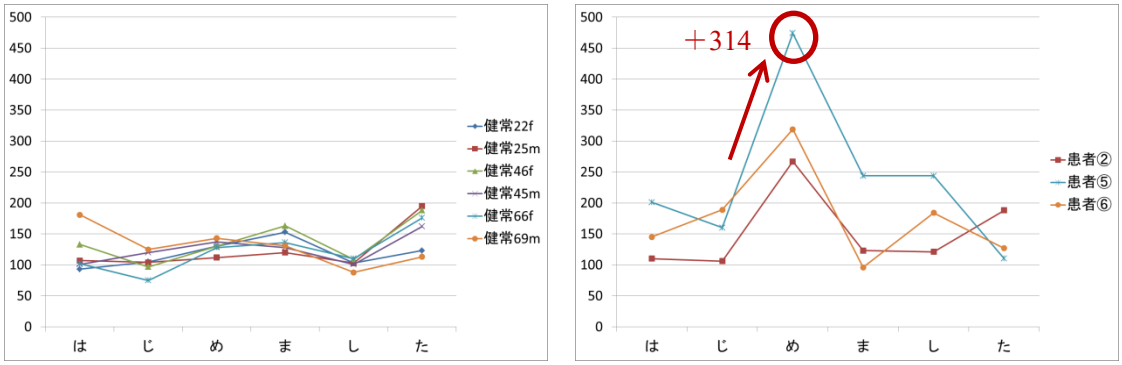


図 4: **健常者と患者の物理的モーラ長「はじめました」(ミリ秒)**

[4]ここでいう＜間＞は「群と群の切れ目」であり、言語の「ポーズ」と同義ではない。言語の「ポーズ」は無音部分の＜間＞のことである。その他、物理的モーラ長が急激に延長しても有音部分の＜間＞になりうる。

例えば図 4 の「はじめました」で、患者⑤は「じ→め」のモーラ持続時間長の差が＋314 ミリ秒で逸脱値となっていた。「め」が顕著に長くなったということは、「め」が群と群の切れ目の＜間＞となりうる。つまり「はじめ」と「ました」が別々の群として認識される。しかし「はじめました」は一つの文節なので、「はじめ」と「ました」は意味と乖離した不適切な群化と言える。この意味と乖離した不適切な群化が、バラバラで異常な印象をもたらしたと考えられる。

6. 結論

リズムの異常性の印象は、＜間＞による不適切な群化によって生じると考えられる。よって聞き手にリズムの異常性を感じさせる時間的要因とは、有音部分および無音部分の不合理な位置での時間延長と言える。そして聞き手にリズムの異常性を感じさせない発話は、少なくとも、表 4 の条件を満たしている必要がある。これは「リズムの異常性がある」発話に対する言語聴覚療法において、評価の客観的指標を示すと共に、より効果的な訓練に繋がるものと期待される。

有音部分	隣接する２モーラの物理的モーラ長が、 統語的・談話的に不合理な位置で急激に増加しないこと。 つまり240ms以上の差あるいは3.0以上の比にならないこと。
無音部分	統語的・談話的に不合理な位置で生じないこと。 合理的な位置であっても、1800msを越えないこと。

表 4: 聞き手がリズムの異常性を感じない発話の必要条件

主要参考文献

氏平明(2008)「特集 正常な発話と逸脱した発話　まえがき」『音声研究』, 12(3), 1-2.

小澤恵美他(2013)『吃音検査法 第２版 解説』　学苑社.

金田一春彦(1967)『日本語音韻の研究』　東京堂出版.

窪薗晴夫(1993)「リズムから見た言語類型論」『言語』, 22(11), 62-69. 大修館.

匂坂芳典(1992)「音声タイミング制御にみられる日本語の特徴」『音声言語医学』, 33(2), 209.

杉藤美代子(1997)「話し言葉のアクセント、イントネーション、リズムとポーズ」
杉藤美代子監修『アクセント・イントネーション・リズムとポーズ』所収, 三省堂.

廣瀬肇他(2001)『言語聴覚士のための運動障害性構音障害』医歯薬出版.

福迫陽子(1983)「麻痺性(運動障害性)構音障害の話しことばの特徴 聴覚印象による評価」
『音声言語医学』, 24, 149-164.

ポール・フレス(1987)「リズムとテンポ」 D.ドイチュ編『音楽の心理学(上)』第 6 章
寺西立年ほか監訳, 182-220. 西村書店.
(Daiana Deutsch,*The Psychology of Music*. *Academic Press*, 1982)

A7

通常小学校在籍の聴覚障害児童の英語分節音産出エラーの特徴
ー摩擦音・破擦音の観察を中心にー

河合裕美（神田外語大学）高山芳樹（東京学芸大学）
kawai@kanda.kuis.ac.jp, yoshiki@u-gakugei.ac.jp

1. はじめに

公立小学校で学ぶ聴覚障害児童は年々増加しており，2020 年の英語教科化を直前に学校の音環境などの合理的配慮の在り方や指導の手立てが急務の課題である。そこで本研究では，聴覚障害児童を含むすべての児童を対象とした英語音声指導法開発の根拠を示すために，通常学級在籍の聴覚障害児童の英語分節音産出のエラー分析を行い，分節音の困難度や特徴を明らかにした。

2. 研究の背景

2.1. 通常学級に在籍する聴覚障害児の英語学習環境

2013 年に学校教育法改正，2016 年には障害者差別解消法が施行され，通常学級での合理的配慮の提供が義務づけられたことにより，聴覚障害特別支援学校に通学する児童数は減少傾向にあるのに対して，公立小学校で学ぶ難聴児童数は増加している（沖津，2016）。英語は 2020 年度から小学校の 5・6 年生において教科となり，「聞く・話す・読む・書く」の 4 技能のすべてを扱うようになるが，英語学習の初期段階においては英語音声に慣れ親しみながら音素認識や音韻認識，さらには分節音に対応する文字が分かる能力（音ー文字一致認識能力）を高めるために，分節音の聴解能力を高めていく必要がある。しかしながら，聴覚障害児童にとっては，日本語モーラでさえ聞き取りづらい上に，連続子音のような英語特有の音声の聴解は困難であるので，当然構音も難しい。これまで難聴児童のための具体的な英語音声指導の手立ては，日本語を介在したものしか存在せず，担当教員や支援員はカタカナ表記をして支援している（村上，2015）。

2.2 子どもの音声言語発達と聴覚障害児童の言語音声知覚・産出の特徴

生まれてからある一定の時期まで，言語構音獲得の順序には普遍性がある。母音が最初に発達し，子音の獲得は，1 歳頃に両唇音から始まって，破裂音と鼻音を獲得し，摩擦音の獲得は比較的遅い。英語母語の子どもの場合，/s, z, v, θ, ð/ などの摩擦音や/tʃ/の破擦音の獲得は 6~8 歳頃までかかる（Templin, 1989）ので，4~5 歳児が/ʃ/を/s/，7 歳を過ぎても/θ/を/f/，5 歳頃までは chair を shair のように代用する現象が見られる（中森，2016）。日本語母語の子どもの場合，小学 1 年生の約 90%が日本語の構音を確立するが，「す・つ・ず・づ」の摩擦音・破擦音の獲得は他の音素の獲得時期と比べるとかなり遅く，6 歳半頃までかかる（高木・安田，1967）。重度聴覚障害児童 1 症例の長期観察においては，構音の獲得順序は聴児とほぼ同じだが，摩擦音や破擦音は 15 才になっても完成しなかった（森・熊井，2017）。小学 4 年生から中学 3 年生の聴覚障害児の語音発音明瞭度（日本語）と聴力レベルの相関性について大規模調査した安東・吉野・志水・板橋（1999）の研究では，発音明瞭度に学年の有意差はなく，破裂音「て」と摩擦音「す」の明瞭度が最も低く，発音のエラーとしては，「が→か」「ば→ぱ」の有声破裂子音の無声化，「す→ふ」の摩擦子音の調音点の誤り，「す→しゅ」の拗音化が顕著に観察されている。日本人の子どもの母音構音については，3 歳児群と 8 歳児群の舌の上下の筋肉の動きには有意差が見られたのに，舌の前後の動きには有意差がなかった。8 歳児でも母音構音において舌の筋肉の前後運動が未熟であることが示唆された（市橋・岡野・近藤・中原・飯沼・田村，2008）。このことは，聴覚障害児童は

もとより，日本人児童にとっては日本語にない英語分節音を構音する場合に，多少なりとも舌の筋肉の発達度が習得の困難度に影響を与えると思われる。Kawai（2017）は，日本のEFL 環境の高学年児童の英語分節音産出能力に関して，子音よりも母音の発音が困難であることを明らかにした。一方，Itabashi（2002）は，130dB を超える最重度難聴児童が /z, ʃ, ʒ/ などの摩擦音を含む音節も正確に発音習得した症例を紹介している。

聾者や聴覚障害者は，低周波域の母音の聴取能力は比較的良好だが，高周波数域の子音では劣るためにその構音能力に影響を与えていることが分かっている（Oyer & Doudna, 1959）。英語音声においては，摩擦音が占める割合は高く，/s/は 4 番目，/z/は 12 番目に頻出度が高い（Tobias, 1959）。これらの分節音は，文法構造において複数形や動詞の時制，所有格など文章の中で重要な役割を担っているので，聴覚障害成人では最もエラーが起こりやすく，誤解が生じやすい（Owens, Benedict, & Schubert, 1972）。動詞の 3 人称単数現在形の語尾の *s* が落ちるなどは，英語母語で軽度から中度難聴児童に起こりやすいエラーである（Elfenbein, Hardin-Jones, & Davis, 1994）。日本語話者聴覚障害者も周波数の高いサ行の聴取が難しいと言われている（加藤・須藤・原島・吉野・江口，1984）。永野（2017）は聴覚障害を持つ生徒にとっては，使用する補聴器の範囲が 125~5000Hz なので/s/が聞こえづらく，さらにサ行の構音は歯裏で息を摩擦させるので十分習得できず，息の量も足りていないと指摘している。

2.4　研究課題：以上のような研究の背景から，以下のような研究課題を設定した。

(1) 聴覚障害児童の英語分節音の産出能力にはどのような特徴があるのか。
(2) 聴覚障害児童の英語分節音の産出能力は，指導によって向上するか。

3.　研究方法

3.1 参加者

参加者は千葉県内公立小学校の 2 校（A・B 校）に在籍，または通級している他の障害を持たない聴覚障害高学年女児 3 名（全員感音性難聴で ID1=中度, ID2=高度重度, ID3=軽度，補聴器は PHONAK Sky-V，デジタルワイヤレス補聴支援システムを使用）の実験群と，A 校の聴児 4 年女児 5 名の統制群である。裸耳聴力は，ID1 と ID2 は周波数が高い聴力が平均聴力より悪く（4000Hz で ID1＝65~70dB，ID2=右 120，左 100dB），ID3 は低い周波数で平均聴力より悪い（125Hz で 60~65dB）（図 1）。本研究実施に際しては，管轄教育委員会や学校長の承諾を経て，参加者保護者と本人に研究協力を要請し，個人情報の取り扱いや本人への配慮事項を伝え，同意を得た上で研究を開始した。

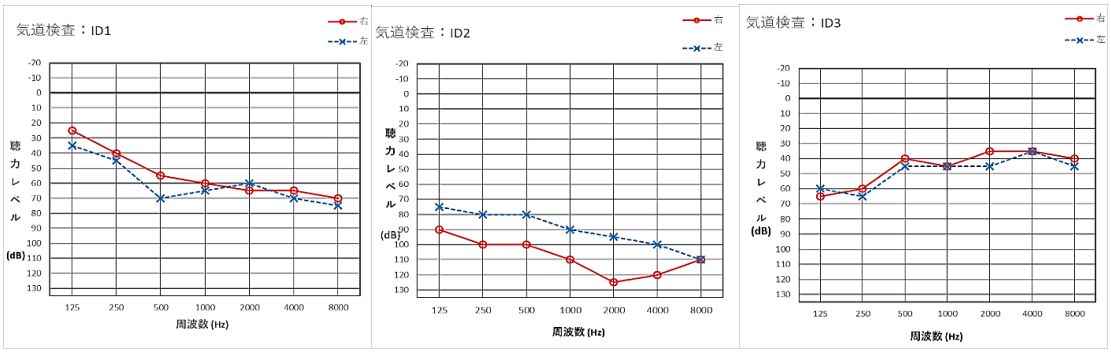


図 1. 聴覚障害児童（実験群）の裸耳聴力（オージオグラム）

3.2 英語音声指導法

参加校のある自治体の公立小学校は，2008 年度より文部科学省教育課程特例校に指定され，１年生から英語を「教科」として導入している。本研究実施の 2018 年度においては，

1~4 年生で週 1 回 20 分の年間 17.5 時間，5・6 年生で 1 回 45 分の年間 50 時間の英語授業が実施されていた。実験群は常に補聴器を装着し，通常教室の英語授業では，担任または ALT 教員がロジャーマイクを装着し，支援員がそばに座って理解が困難な場合は書いて示す（日本語カタカナ表記）などの支援を受けていた。実験群は個別取り出しの自立支援活動として，子音を中心に音素認識能力を高め，常に指導教員の口形や口周辺の筋肉の動きを「見る」指導を約 4 ヶ月間受けた。子音の明示的な指導順序は，破裂音や両唇音鼻音の判別から始め，次第に摩擦音・破擦音を導入した。

3.4 英語分節音産出テスト

英語母語子どもの音声処理システムのモデル（Stackhouse & Wells, 1997）を応用したテストを Kawai（2017）が改編したものに，さらに摩擦音や破擦音を頭音に持つ単語を追加し，現実単語（1~3 音節単語で計 212 分節音）と非単語（1~3 音節で計 182 分節音）の産出テストを指導の事前事後に実施した。表 1 が示すように，非単語は基本的に現実単語の母音を置換して意味表象を遮断した。防音施工の聞こえ教室で，モデル話者となる男性英語母語話者の顔が見えるように対面して座ってもらい，モデル話者の発音を真似してもらった。対面距離は約 1 メートルである。統制群には同じモデル話者の発音を録画し，ヘッドセットを装着し PC 上で同じテストを一度受けてもらった。産出データは Roland R-05 録音機（ver. 1.03 WAV-24bit）を使って録音した。録音機と受験者の距離は約 30 センチである。録音データは，評価者 3 名（音声教育を専門とする大学教員 2 名と，アメリカ・カナダ国籍の A 校の ALT 教員）が，産出された分節音全てをモデル話者と同じか違うか（1/0）で評価し，さらに，エラー分析と Praat による観察を行った。

表 1．テスト単語例（1 音節の場合）

現実単語	IPA記号	非単語	分節音数
fish	/fɪʃ/	/fɛʃ/	3
cat	/kæt/	/kɪt/	3
sheep	/ʃip/	/ʃɛp/	3
zoo	/zu/	/zi/	2
gym	/dʒɪm/	/dʒɛm/	3
socks	/sɑks/	/sæps/	4

Note. 非単語は，参加児童にとって未知と考えられる単語である。

4．結果・考察

4.1 現実単語・非単語産出テスト記述統計

表 2. 現実単語（全 212 音素）と非単語（全 182 音素）の記述統計（評価者 A）

	児童ID	学年	現実				非単語			
			事前		事後		事前		事後	
			M (*SD*)	得点 (%)	*M* (*SD*)	得点 (%)	*M* (*SD*)	得点 (%)	*M* (*SD*)	得点 (%)
実験群	1	4	158.3 (11.930)	153 (72.2%)	174.0 (2.646)	173 (81.6%)	136.0 (4.359)	134 (73.6%)	143.7 (6.028)	143 (78.6%)
	2	6		150 (70.8%)		172 (81.1%)		133 (73.1%)		138 (75.8%)
	3	6		172 (81.1%)		177 (83.5%)		141 (77.5%)		150 (82.4%)
統制群	9	4	176.8 (21.925)	171 (80.7%)			145.4 (19.756)	113 (62.1%)		
	10	4		198 (93.4%)				149 (81.9%)		
	11	4		194 (91.5%)				156 (85.7%)		
	12	4		178 (83.9%)				165 (90.7%)		
	13	4		143 (67.5%)				144 (79.1%)		

Note. 現実単語の評価者間信頼性は事前α = .935，事後α = .857，非単語は事前 α = .899，事後 α = .899であった。

3 名の評価者間信頼性は，十分な信頼性が得られた。表 2 に評価者 A が評価した結果を示す。現実単語・非単語ともに事前の実験群の評価平均は統制群平均より低いが，事後の平均点は事前より向上し，実験群の 3 名とも事前より事後の点数が高かった。

4.2 エラー分析

次に産出された英語分節音にどのようなエラーがあるのかを観察した。実験群が2回ずつ産出した分節音を評価者全員がエラーと判定した分節音（下線）と，実験群の3名ともがエラーと評価された分節音（下線）を単語レベルで事前・事後で表3にまとめた。

表3．現実単語（上段）・非単語（下段）テストにおけるエラー分節音（箇所）

児童ID	事前 評価者3名全員がエラーと判定した分節音（単語に下線）	事後 評価者3名全員がエラーと判定した分節音（単語に下線）
ID1（中度）	sponge, sandwich, hospital, money, umbrella, leaf, flower, computer, train, hamburger, zoo	sponge, sandwich, hospital, umbrella, leaf, flower, computer, socks
	/pləʊt/ /vɪn//fɛʃ/ /bɹɪʃ//spéɪdɪ/ /tɹɛ́ktɪ/ /slɔ́pə//gɛ́tɔ/ /spʌgɛ́tə/ /ǽlɪfɒnt/ /bǽtəflə́ʊ//zi/ /sæps/	/fɛʃ//spéɪdɪ//bǽtəflə/spʌgɛ́tə/ /ǽlɪfɒnt//zi/
ID2（重度）	sponge, sandwich, hospital, duck, money, umbrella, leaf, kitchen, kangaroo, flower, computer, train, toilet, hamburger, socks, zoo, cat, salad	sponge, sandwich, umbrella, kangaroo, toilet, hamburger, socks, salad
	/pləʊt/ /vɪn/ /bɹɪʃ/ /spéɪdɪ/ /tɹɛ́ktɪ/ /dʒʌ́lə/ /slɔ́pə/ /spʌgɛ́tə/ /pɛ́ɹəʃɪt/ /ǽlɪfɒnt/ /zi/ /ʃɛp/ /kɪt/ /sæps/	/bɹɪʃ/ /spéɪdɪ/ /tɹɛ́ktɪ/ /slɔ́pə/ /spʌgɛ́tə/ /pɛ́ɹəʃɪt/ /ǽlɪfɒnt/ /zi/ /ʃɛp/ /sæps/
ID3（軽度）	sponge, computer, train	hospital, umbrella, computer, socks
	/vɪn/ /dʒʌ́lə/ /slɔ́pə/ /gɛ́tɔ/ /bǽtəflə́ʊ/ /ǽlɪfɒnt/ /dʒɛm/ /sæps/	/tɹɛ́ktɪ/ /dʒʌ́lə/ /gɛ́tɔ/ /bǽtəflə́ʊ/ /spʌgɛ́tə/ /pɛ́ɹəʃɪt/
3名とも間違い	knife, hamburger	knife, hamburger
	/glɛv/ /pɪdʒɔ́mɪz/	/glɛv/ /pɪdʒɔ́mɪz/

さらに実験群と統制群のエラーの特徴に違いがあるのかを観察した。エラーと評価された分節音の割合を種類ごとに示したものが表4である。事前で見ると，実験群・統制群ともに最もエラーが多いのが/ɹ//l//w/などの側音・半母音である。/s//z//f/などの摩擦音は，実験群だけでなく統制群でもエラー率が高い。両群ともに，摩擦音のエラーは現実単語で高く，非単語ではエラー率がずっと低いことは興味深い。ところが摩擦音に近い/tʃ//dʒ/などの破擦音は，実験群は現実単語も非単語も30%以上のエラー率である一方で，統制群は10%台である。破裂音や鼻音については，実験群が事前でエラー率が統制群より高いが，事後で両者の差が縮まっており，実験群の鼻音非単語の事後ではエラーは観察されなかった。統制群では，側音・半母音の次に，母音のエラー率が現実単語も非単語も高く，Kawai（2017）の研究を支持する結果となった。実験群の事前・事後においては，破裂音非単語ではエラー率が増え，摩擦音非単語では横ばいであった。それ以外の分節音においては，事後が事前よりエラー率が低かったことは，個別指導の成果と考える。実験群事後において，最もエラー率が高かったのは側音・半母音で，次いで破擦音であった。実際の個別指導では，子音に着目し，破裂音⇒鼻音⇒摩擦音・破擦音の順序で指導していったが，破擦音の指導時間は十分ではなく，日本人英語学習者が一般的に苦手としている側音・半母音・母音の指導まで及ぶことができなかった。それにも関わらず，それらの分節音産出のエラー率が下がっていたことは特筆すべき点である。また，聴覚障害児と聴児の両グループともに摩擦

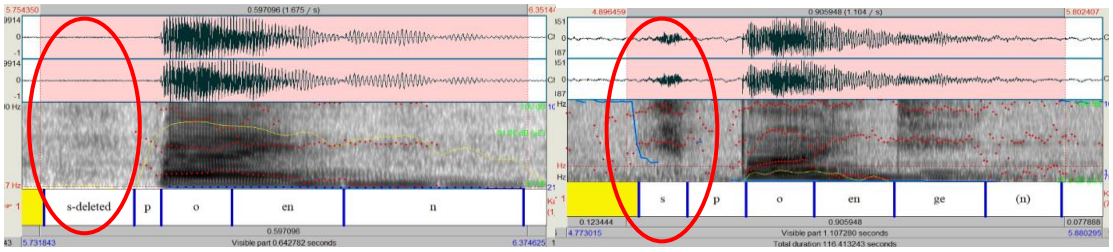
音のエラー率が高かったことは，聴覚障害の有無に関わらず周波数の高い摩擦音の獲得が児童にとって容易ではないことを示唆しており，小学校環境で指導する上で特に考慮すべき点であると考える。

表4．英語分節音産出エラー率（%）

	母音		破裂音		鼻音		破擦音		摩擦音		側音・半母音	
	現実単語	非単語	現実単語	非単語	現実単語	非単語	現実単語	非単語	現実単語	非単語	現実単語	非単語
実験群事前	22.1	26.5	13.0	10.7	9.7	25.0	31.1	38.5	41.7	5.6	60.3	40.7
実験群事後	11.7	22.1	9.4	16.0	6.9	0.0	27.8	32.3	20.8	5.6	51.3	25.9
統制群（聴児）	20.8	30.3	5.2	8.4	4.2	15.0	10.7	16.3	37.5	3.3	38.5	26.7

4.3　Praat上のエラー観察

実験群の英語分節音産出は，事前より事後においてエラーが減った。加えて，評価者らの聴覚的印象は，3名とも事前より事後の産出音声の方がはっきりと発声されており，多少なりとも「上手になった」とのことであった。この評価者の主観的評価の根拠を裏付けるため，Praat上で分節音ごとに区切り，波形やスペクトログラムの観察を試みた。本稿では聴覚障害者が知覚・産出が困難である傾向が強い摩擦音を取り上げる。“sponge”の頭音/s/では，事前でID2（高度重度女児）の/s/は全く産出されていない状態であった。しかしながら，事後において/s/の波形やスペクトログラムをはっきりと観察することができる（図2参照）。当該箇所の評価は，評価者3名中2名が正答と評価していた。/z/については，現実単語の“zoo”の場合は，ID1は事前の/z/が別の分節音で産出されていたが，事後では/z/が産出されていた。一方，非単語/zi/の場合は，事前・事後とも評価者全員がID1・ID2は誤答と判定したが，聴力が軽度のID3は前述の“sponge”も/zi/も評価者3人全員が事前・事後ともに正答と判定した。/z/は個別指導中にID1もID2も/dʒ/との判別が大変困難で，指導に時間をかけた分節音である。ID2のスペクトログラムでは，1回目よりも2回目で近い音にしようという試みが確認されたものの，評価者にはエラーと判別された。意味表象を伴わない非単語は，聴力への依存度が大きいため，重度の聴覚障害児童にとっては補聴器を装用していても産出が困難であったと思われる。



ID2 事前“sponge”（頭音の/s/音の呼気が全くない）　ID2 事後“sponge”（頭音の/s/音の呼気が認められる）

図2．Praatで観察したID2（高度重度女児）の“*sponge*”スペクトログラムと波形

これらの結果から，聴覚障害児童に音声指導をする際には，意味表象有の現実単語を使ってコンテクストのある音声指導をする方が効果的であると思われる。また，ID1・ID2のエラーとID3では明らかに困難さに差があると思われ，軽度と中度の間で弁別閾値が存在し，産出に影響を与えていることを示唆する結果となった。

５．結論・今後の課題

聴覚障害児童の英語分節音エラーは摩擦音，破擦音，側音・半母音が多く，摩擦音，側音・半母音は聴児でもエラーが多い。通常学級のクラスサイズや騒音を考慮すると，聴児でもそれらの分節音は聴き取りづらく，母音と同様に，特に明示的な構音指導が必要である。本研究では，破擦音の先行研究が十分になく，今後さらなる検証が必要である。

謝辞.

本研究は博報財団 2018 年度第 13 回児童教育実践についての研究助成に採択された研究の一部を加筆・修正したものです。

参考文献

安東孝治・吉野公喜・志水康雄・板橋安人 (1999).「聴覚障害児における語音明瞭度，発音明瞭度並びに聴力レベルの相互関連性について」『特殊教育学研究』*36*(4)，49−57.

市橋豊雄・岡野哲・近藤亜子・中原弘美・飯沼光生・田村康夫 (2008).「持続母音およびVCV音節語後続母音の周波数解析からみた小児の構音発達について」『小児歯科学雑誌』*46*(5)，585−601.

沖津卓二 (2016).「普通学校における難聴児への対応」第 11 回日本耳鼻咽喉科学会，*37*(3)，241−245.

加藤靖佳・須藤正彦・原島恒夫・吉野公喜・江口実美 (1984).「聴覚障害者における無声摩擦音/s/-/ʃ/の識別の研究」*Audiology Japan*，*27*(5)，621−622.

高木俊一郎・安田章子 (1967).「正常幼児 (3~6才)の構音能力」『小児保健研究』第25巻，第1号，23−28.

中森誉之 (2016).『外国語音声の認知メカニズムー聴覚・視覚・触覚からの信号ー』東京：開拓社.

永野哲郎 (2017).『聴覚障害児の発音・発語指導ーできることを，できるところからー』東京：ジアース教育新社.

村上理恵子 (2015).『通常の学級におけるきこえにくい子どもの外国語活動ーきこえにくい子どもの困り感から考える手立てや工夫ー』平成 27 年度千葉県長期研修生研究報告書.

森つくり・熊井正之 (2017).「重度難聴児の構音能力の長期経過ー補聴器装用例についてー」*Audiology Japan*, *60*, 210−218.

Elfenbein, J. L., Hardin-Jones, M. A., & Davis, J. M. (1994). “Oral communication skills of children who are hard of hearing.” *Journal of Speech and Hearing Research*, *37*, 216–226.

Itabashi, Y. (2002). “Improvements of speech production skills in a child after prolonged cochlear implant use.” *Paper presented in The 8th Asia-Pacific Congress on Deafness, August 3–6, Taipei, Taiwan, Proceedings in CD-ROM.*

Kawai, H. (2017). *A Study of the English speech processing system in young Japanese EFL learners and changes in their awareness through explicit sound instruction*（青山学院大学大学院文学研究科英米文学専攻博士学位論文）.

Owens, E., Benedict, M., & Schubert, E. D. (1972). “Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairment.” *Journal of Speech and Hearing Research*, *15*, 308–322.

Oyer, H. J., & Doudna, M. (1959). “Structural analysis of word responses made by hard of hearing subjects on a discrimination test.” *A.M.A. Archives of Otolaryngology*, *70*, 357−64.

Stackhouse, J., & Wells, B. (1997). *Children’s speech and literacy difficulty 1*: *A psycholinguistic framework*. London, UK: Whurr Publishers.

Stelmachowicz, P. G., Nishi, K., Choi, S., Lewis, D. E., Hoover, B. M., & Dierking, D. (2008). “Effects of stimulus bandwidth on the imitation of English fricatives by normal-hearing children.” *Journal of Speech Language and Hearing Research*, *51*(5), 1369–1380.

Templin, M. (1957). *Certain language skills in children: their development and interrelationships*. University of Minnesota Institute of Child Welfare Monograph Series 26. Minneapolis, MN: University of Minnesota Press.

Tobias, J. V. (1959). “Relative occurrence of phonemes in American English.” *Journal of the Acoustical Society of America*, *31,* 631.