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.

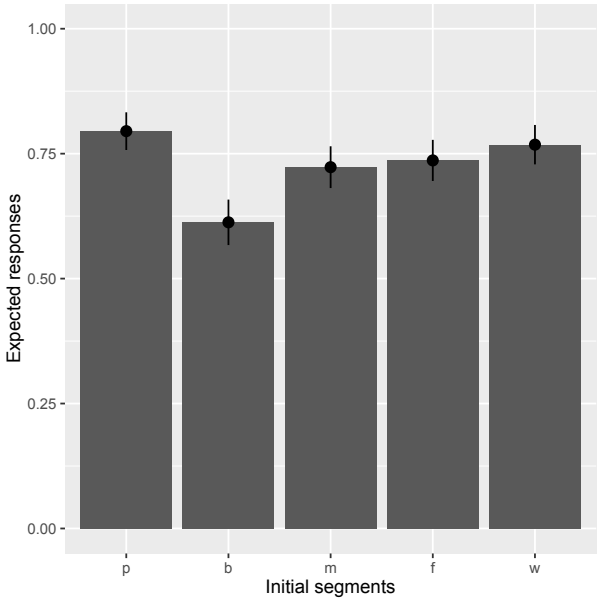

図 1: 唇音を含む刺激が選ばれた割合
（エラーバーは 95%信頼区間）

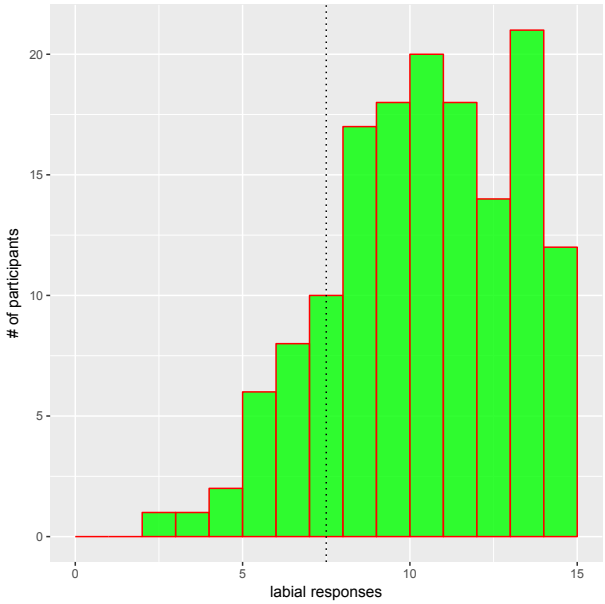

図 2: 「唇音反応数」のヒストグラム
（y 軸は被験者数）

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．

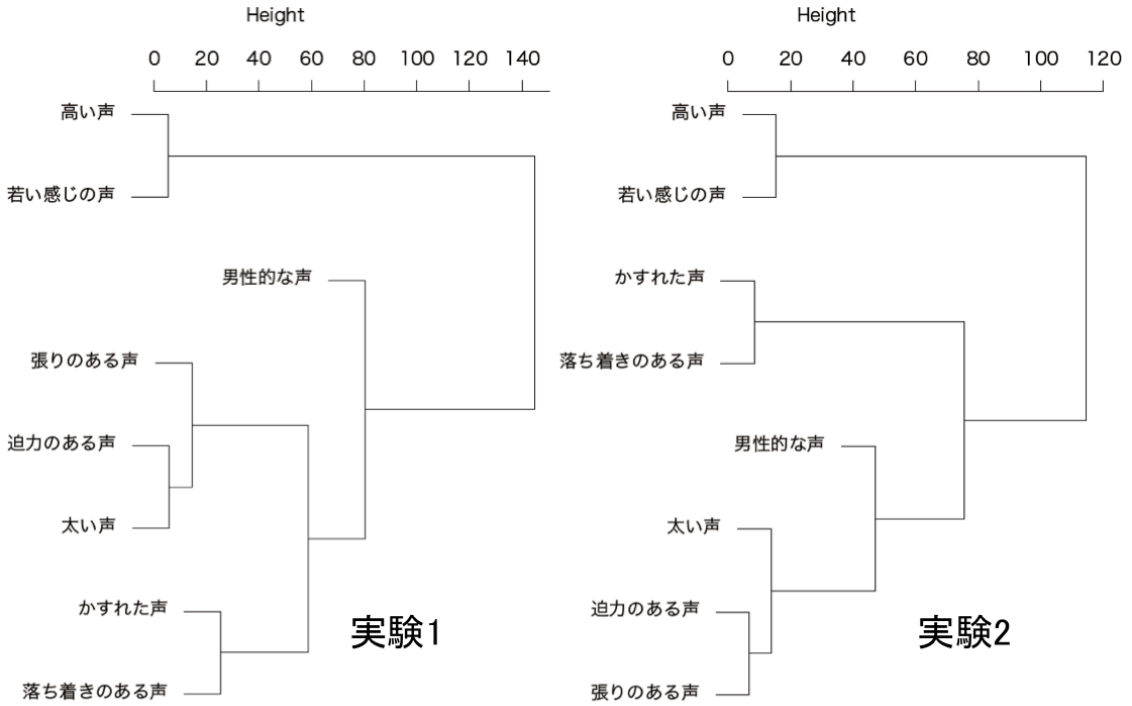

図 1: 表現語を対象としたクラスター分析の結果．（左）実験 1．（右）実験 2．

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．

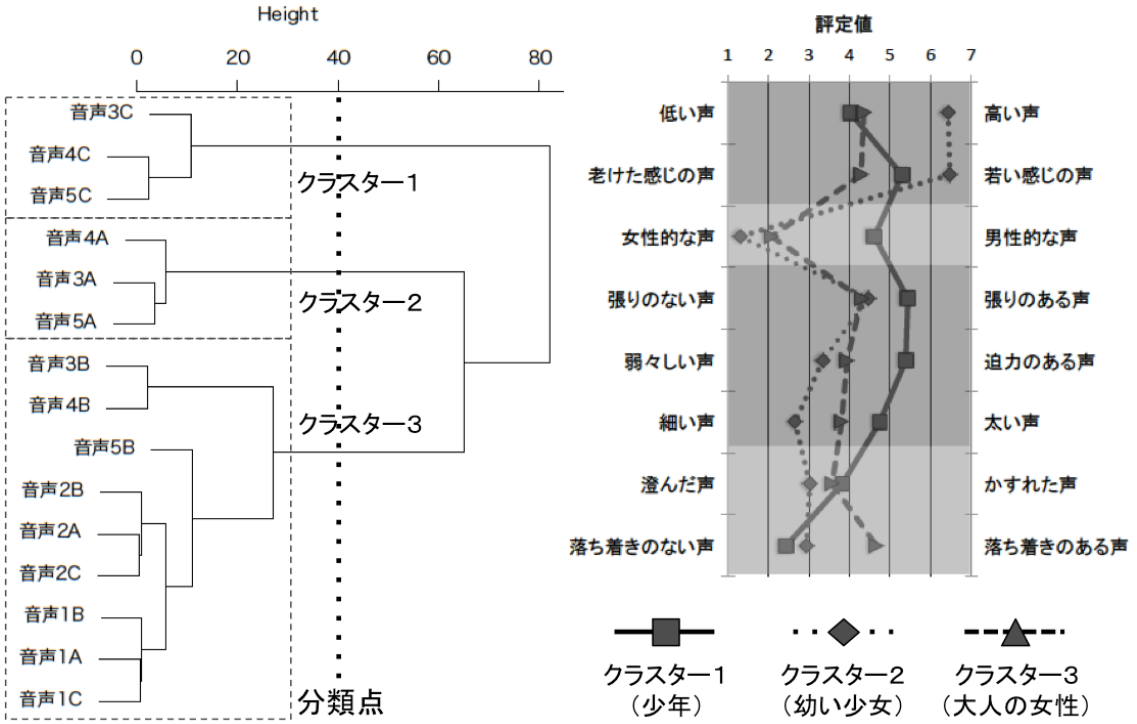

図 2: 実験１の音声の分類．（左）クラスター分析により得られたデンドログラム．（右）クラスターごとの評定値の平均．

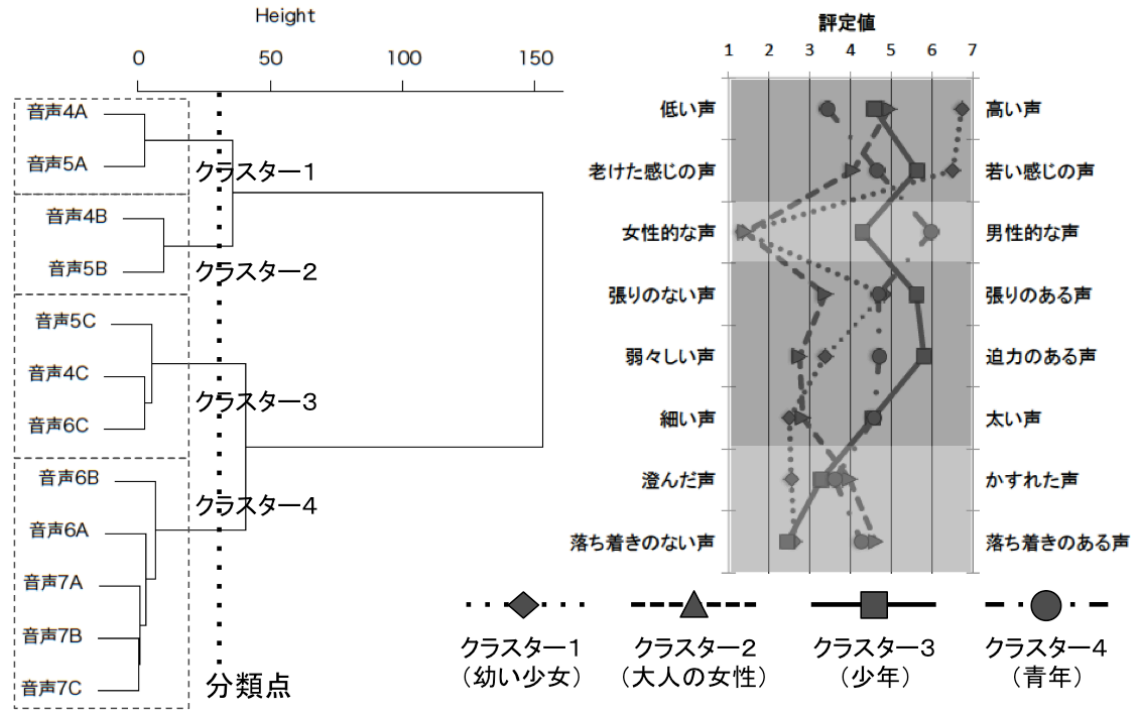

図 3: 実験２の音声の分類．（左）クラスター分析により得られたデンドログラム．（右）クラスターごとの評定値の平均．

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．

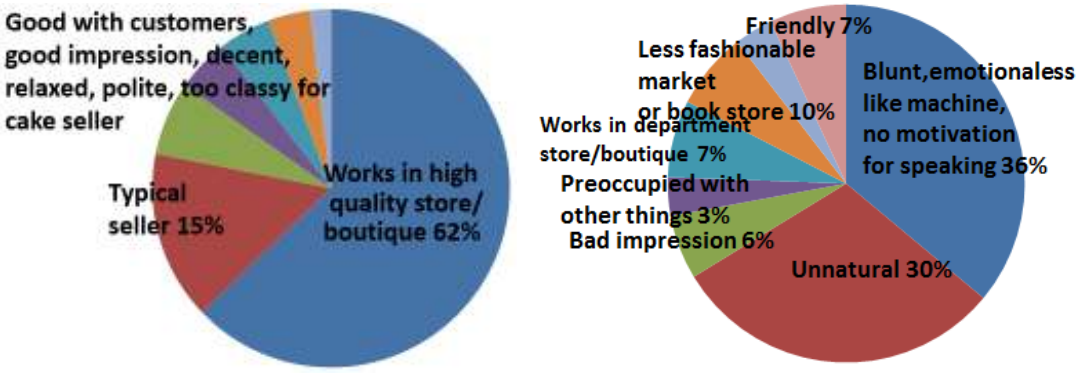

図 1:ケーキの売り子の声としての PS(左)・PU(右)に対する日本語母語話者の評価(自由記述).

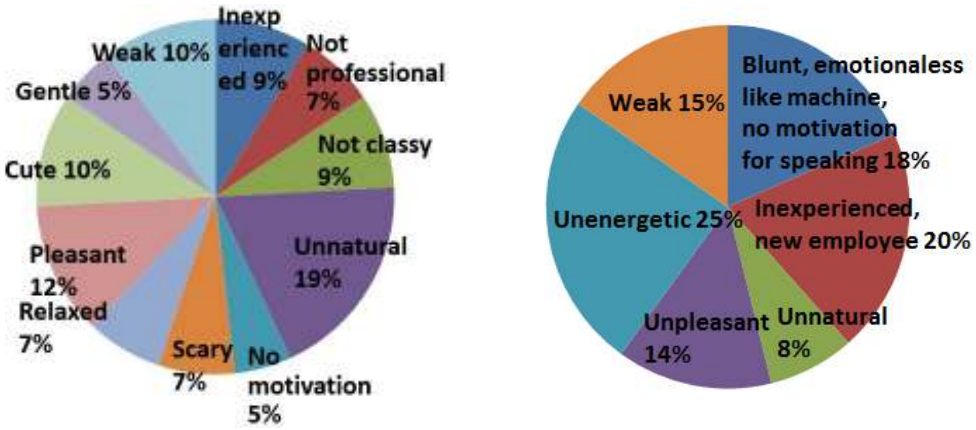

図 2:ケーキの売り子の声としての QS(左)・QU(右)に対する日本語母語話者の評価(自由記述).

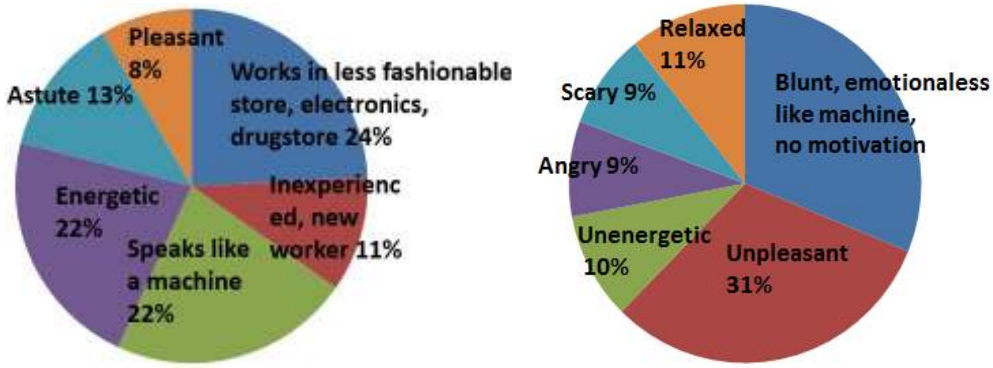

図 3:ケーキの売り子の声としての RS(左)・RU(右)に対する日本語母語話者の評価(自由記述).

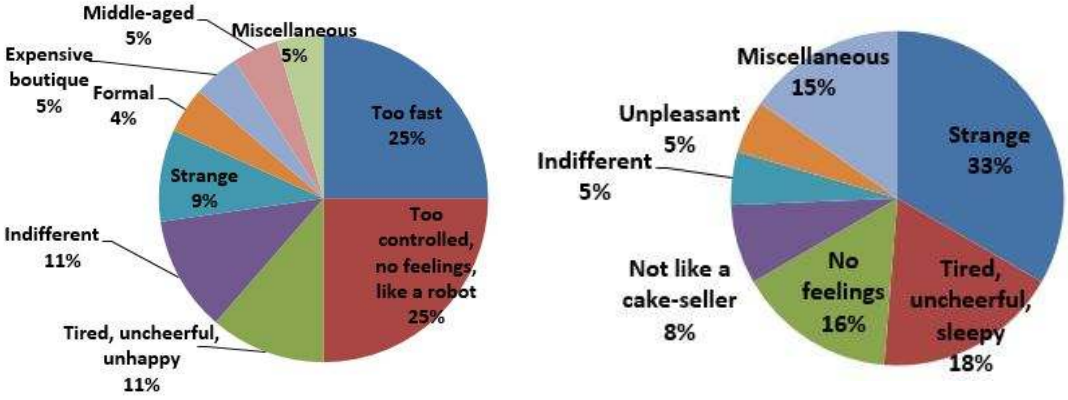

図 4:ケーキの売り子の声としての PS(左)・PU(右)に対する中国語母語話者の評価(自由記述).

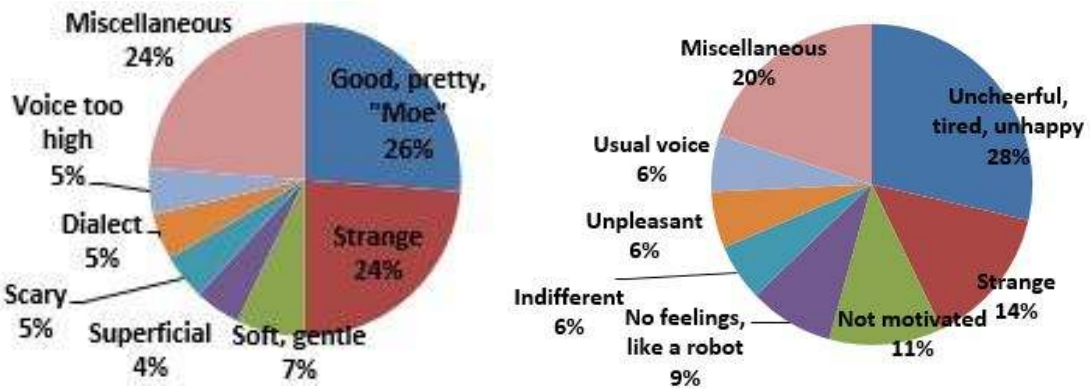

図 5:ケーキの売り子の声としての QS(左)・QU(右)に対する中国語母語話者の評価(自由記述).

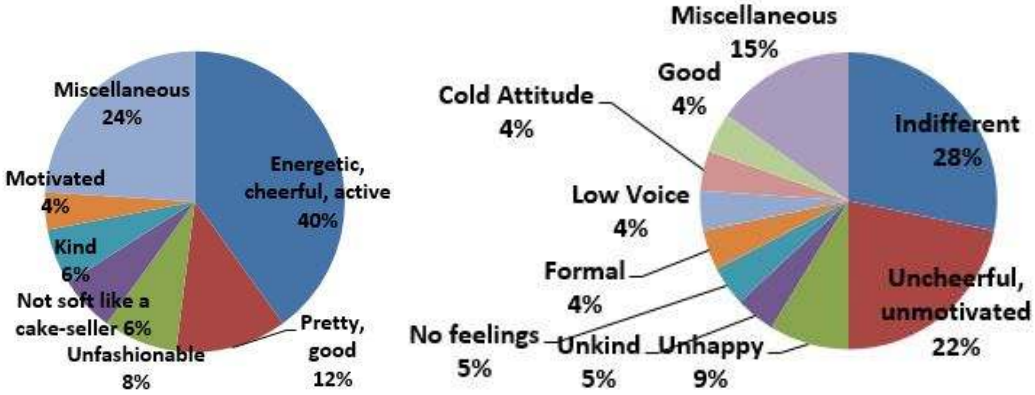

図 6:ケーキの売り子の声としての RS(左)・RU(右)に対する中国語母語話者の評価(自由記述).

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：

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

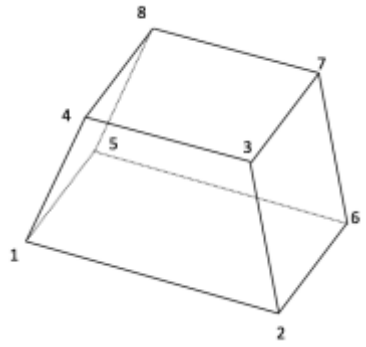

図 1：6 面体要素

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．

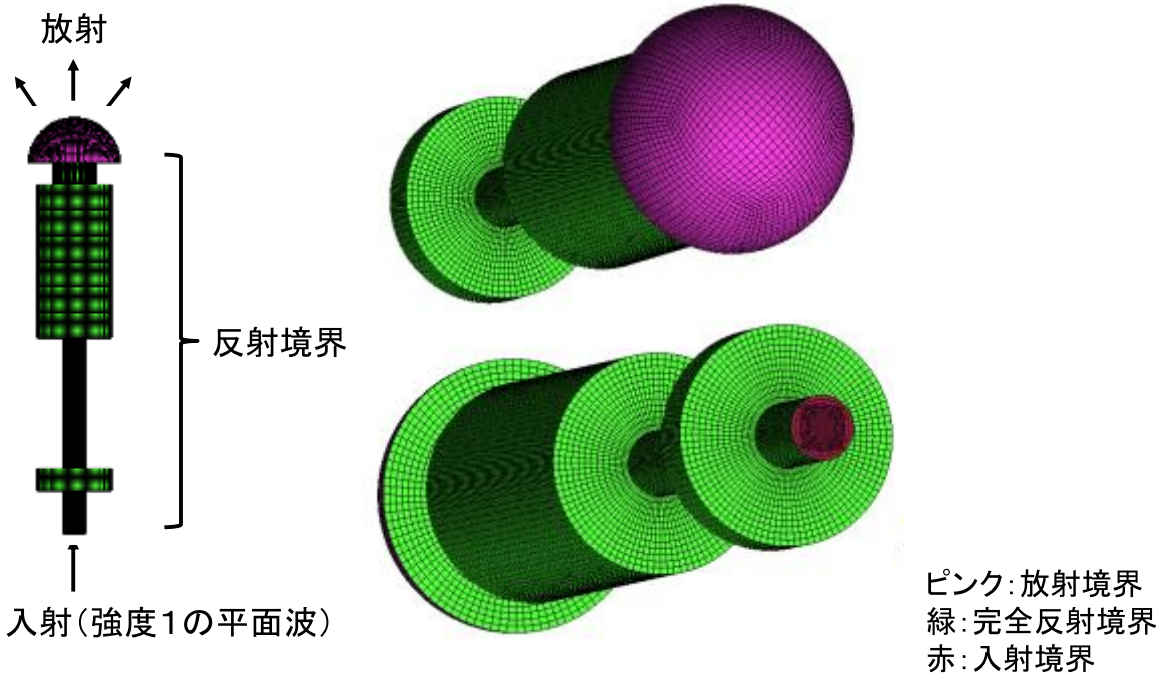

図 2：解析条件．「お」の管．

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）．

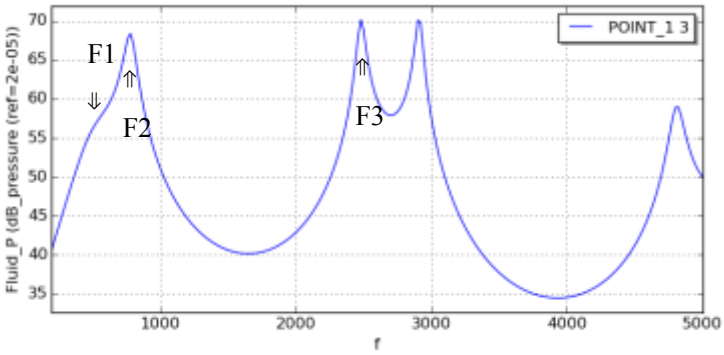

図 3(a)「お」の FEM 解析結果

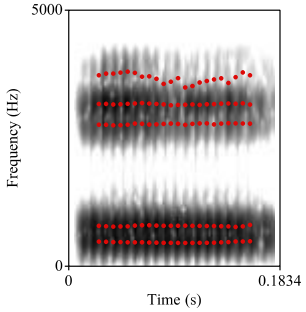

図 3(b) 自然発話の「お」

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．

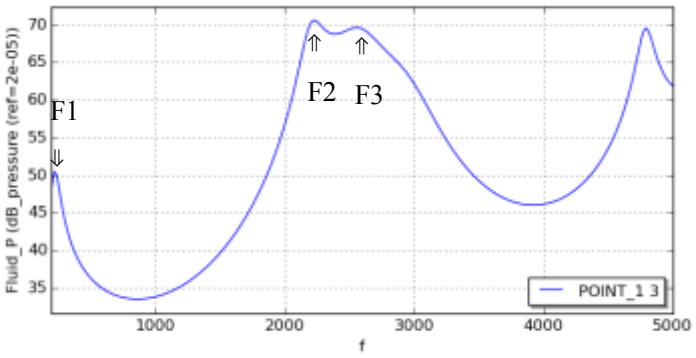

図 4(a)「い」の FEM 解析結果

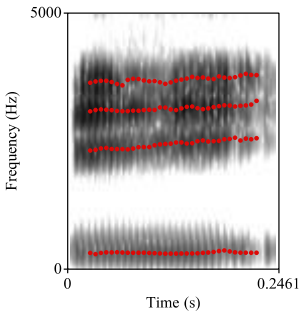

図 4(b) 自然発話の「い」

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．

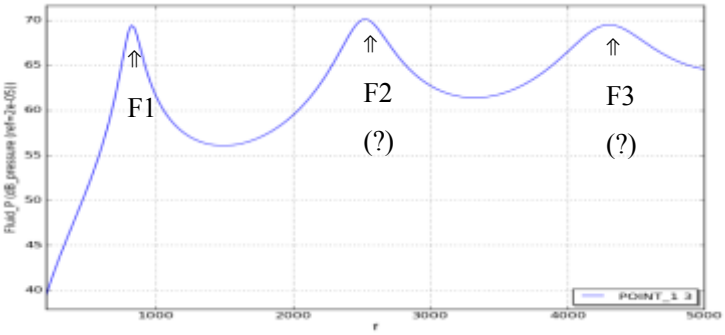

図 5(a) 「あ」の FEM 解析結果

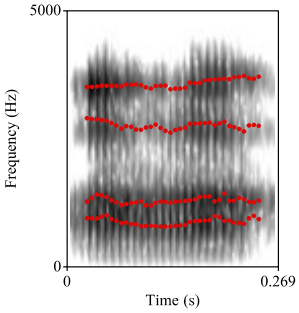

図 5(a) 自然発話の「あ」

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。

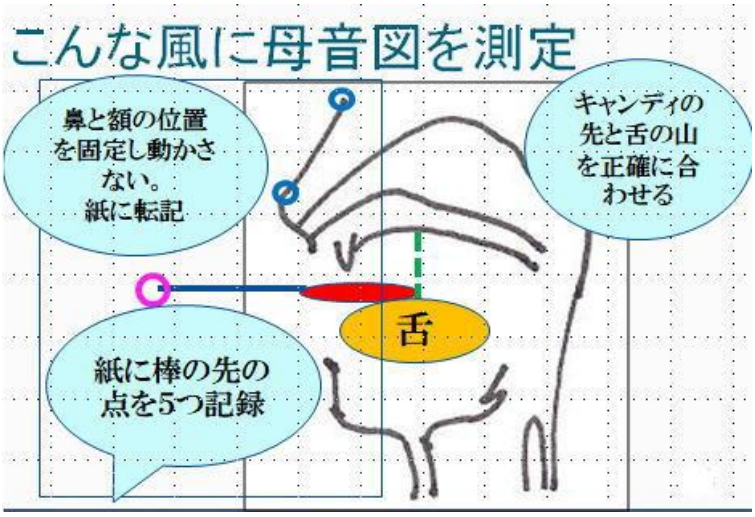

図 1: 母音図測定の様子

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

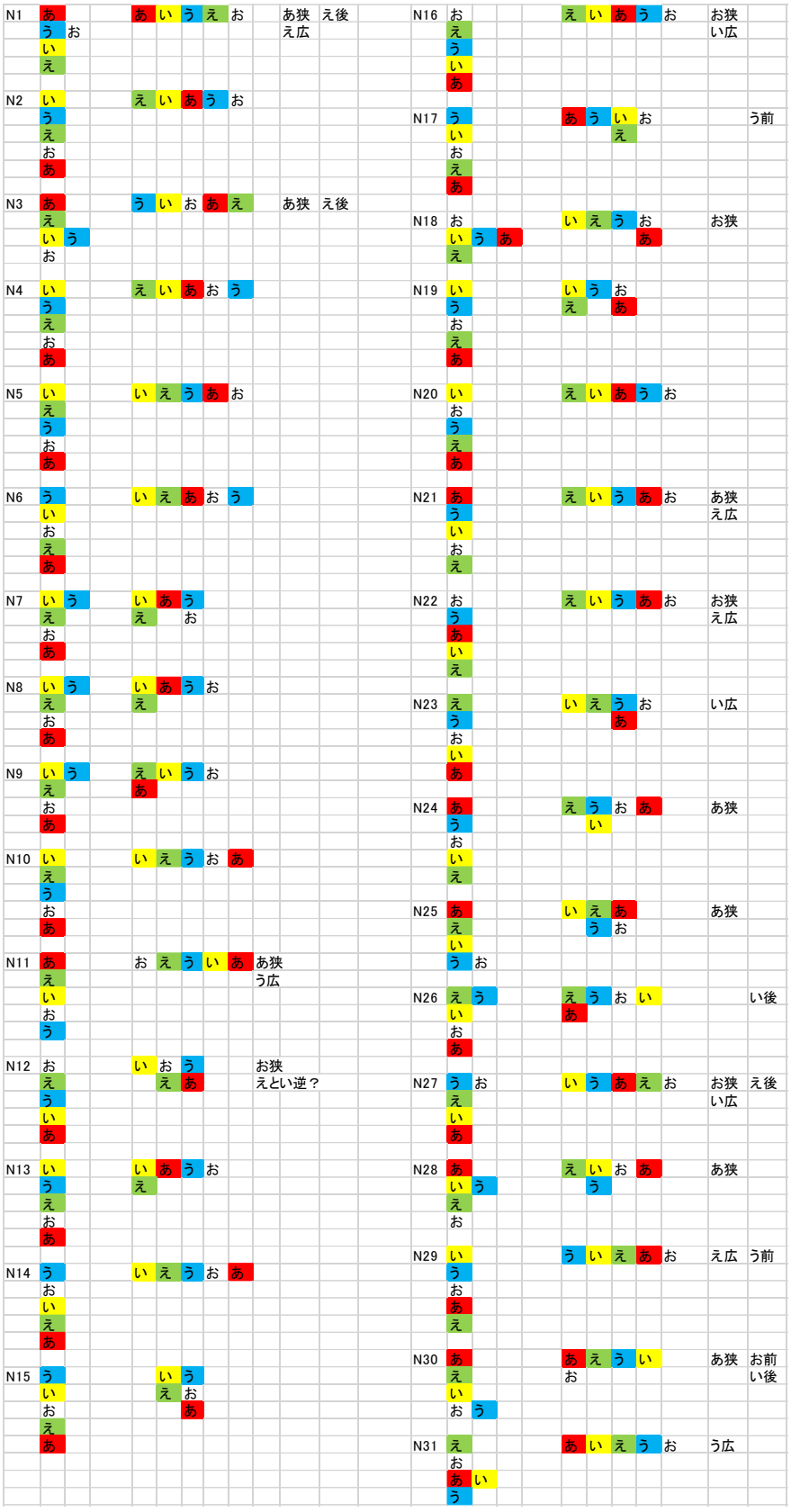

図 2: 各被験者のデータ-2

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.

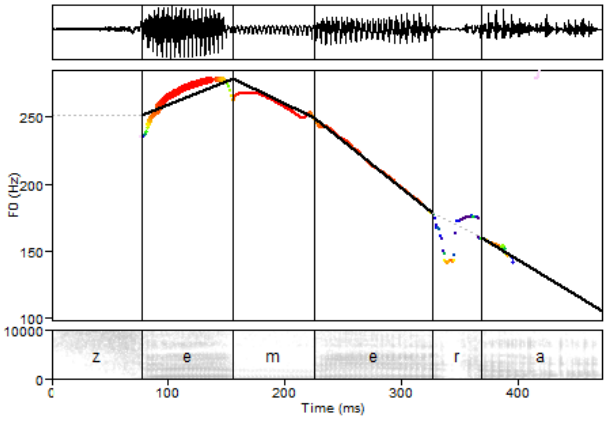

Figure 1: Example of output, applied to nonword ***zémera***

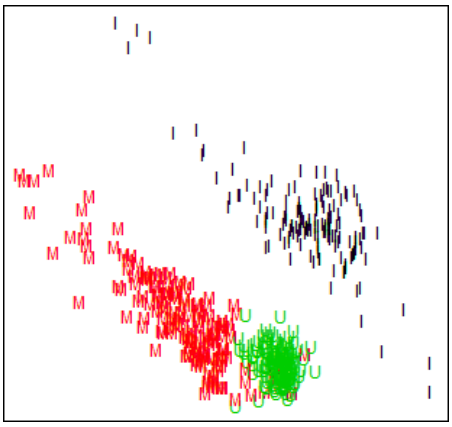

Figure 2: Multidimensional scaling results

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.