

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

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

1. Introduction

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

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

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

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

2. Method

2.1. Participants

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

2.2. Materials

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

Table 1: An example of a stimulus *gaoxing* ‘happy’ /kəʊeiŋ/ and its prime and target conditions.

	Prime (auditory)			Target	
	Canonical	Variant	Control		
Exp. 1	[kəʊeiŋ]	[kəʊeiŋ]	<i>kəfei</i> ‘coffee’	高興	(visual)
Exp. 2	[kəʊeiŋ]	[kəʊeiŋ]	<i>kəfei</i> ‘coffee’	Canonical: Variant:	[kəʊeiŋ] [kəʊeiŋ] (auditory)

2.3. Procedure

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

3. Results

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

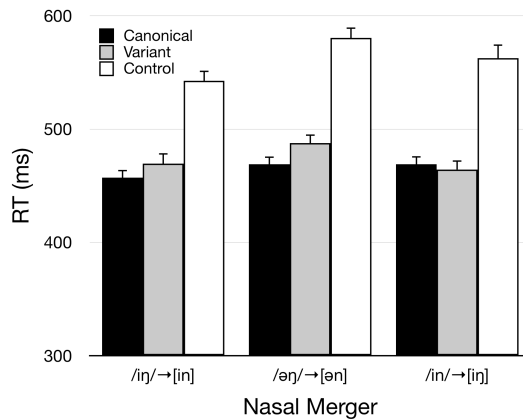


Figure 1: RTs of the three priming conditions in Experiment 1.

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

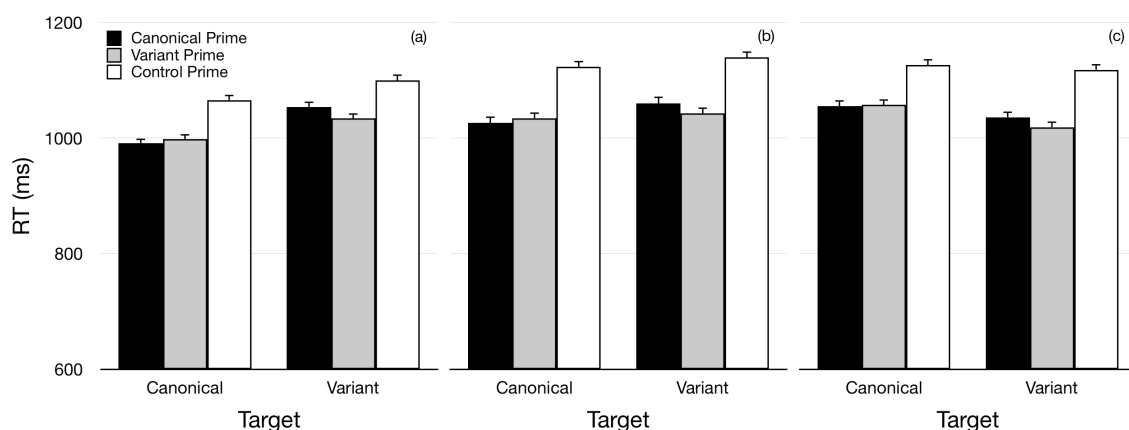


Figure 2: RTs for (a) /iŋ/→[in], (b) /əŋ/→[ən], and (c) /in/→[iŋ] in Experiment 2.

4. Discussion

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

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

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

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

5. Conclusion

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

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

References

- Fon, J., Hung, J.-m., Huang, Y.-H., & Hsu, H.-j. (2011) "Dialectal variations on syllable-final nasal mergers in Taiwan Mandarin", *Language and Linguistics* 12(2), 273-311.
- Gaskell, M. G. & Marslen-Wilson, W. D. (1998) "Mechanisms of phonological inference in speech perception", *Journal of Experimental Psychology: Human Perception and Performance* 24(2), 380-396.
- Gow, D. W. (2002) "Does English coronal place assimilation create lexical ambiguity?", *Journal of Experimental Psychology: Human Perception and Performance* 28(1), 163-179.
- Johnson, K. (2004) "Massive reduction in conversational American English." In K. Yoneyama and K. Maekawa (eds.) *Spontaneous Speech: Data and Analysis*. (pp. 29-54). Tokyo: The National International Institute for Japanese Language.
- Pitt, M. A. (2009) "The strength and time course of lexical activation of pronunciation variants", *Journal of Experimental Psychology: Human Perception and Performance* 35(3), 896-910.
- Ranbom, L. J., & Connine, C. M. (2007) "Lexical representation of phonological variation in spoken word recognition", *Journal of Memory and Language* 57(2), 273-298.
- Snoeren, N. D., Seguí, J., & Hallé, P. A. (2008) "On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French", *Cognition* 108(2), 512-521.
- Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014) "The socially weighted encoding of spoken words: a dual-route approach to speech perception", *Frontiers in Psychology* 4.
- Sumner, M. & Samuel, A. G. (2009). "The effect of experience on the perception and representation of dialect variants", *Journal of Memory and Language* 60(4), 487-501.
- Wu, J.-T., Yang, F.-L., & Lin, W.-C. (2013) "Beyond phonology matters in character recognition", *Chinese Journal of Psychology* 55(3), 289-318.

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

王 鳳翔 (広島大学大学院)
dl64478@hiroshima-u.ac.jp

1. はじめに

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

2. 先行研究及び先行研究の問題点

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

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

3. 生成調査

3.1. 調査対象

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

3.2. 調査手順

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

3.3. 調査用語

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

4. 生成調査—結果の分析と考察

動詞打ち消し形の正答率について、その結果は、以下のようになる。

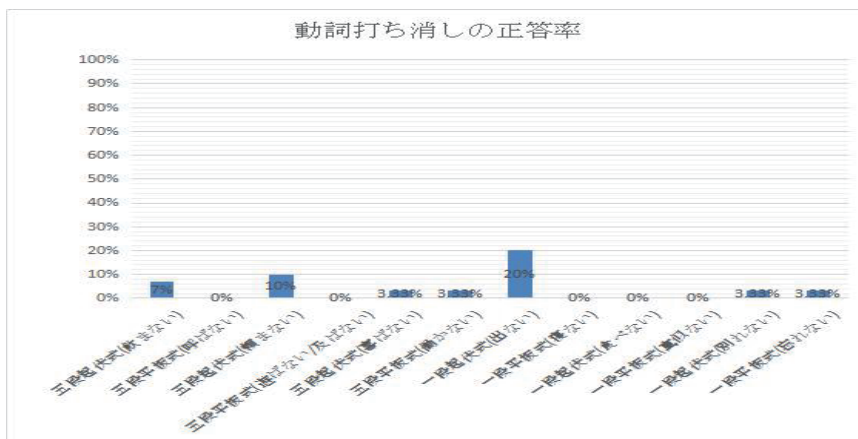


図 1: 動詞打ち消し形の正答率

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

表 1: -2 型アクセントで動詞打消しを発音する話者の比率(%)

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

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

4.1. 形容詞（○○ない）との関わり

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

表 2: 各教材の形容詞と動詞打ち消しの学習の順番

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

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

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

4.2. プロミネンスとの関わり

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

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

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

1) 日本語

Q:あなたは林檎を食べたいですか?

A:(私は林檎を)食べたくないです。

2) ドイツ語

Q: Möchtest du einen Apfel essen?

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

3) 英語

Q: Do you want to eat an apple?

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

4) 中国語

Q: 你想吃苹果吗?

A: 不, 我不想吃(苹果)。

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

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

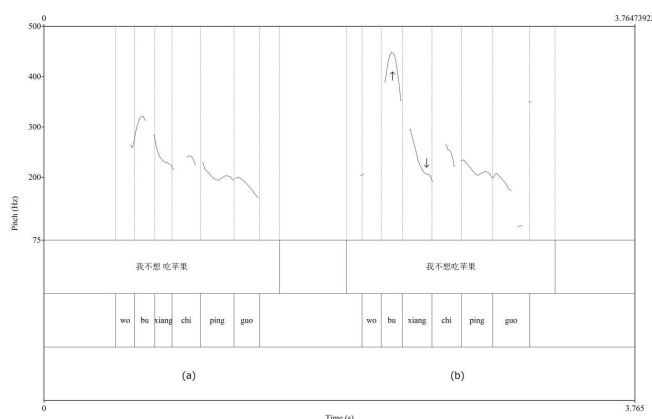


図 2: 我不想吃苹果[uo²¹⁴ pu⁵¹ əiɑŋ²¹⁴ t̚^hh⁵⁵ p^hiŋ³⁵ kuo²¹⁴]

図 2 では、(a)は感情を込めず棒読みで「我不想吃苹果」を発音した。特にプロミネンスを置かない中立的な発話である。(b)は「不」のところにプロミネンスを置いた発話である。

(a)の発話と比べ、(b)は「↑」で示した部分が高く、「↓」で示した部分は(a)より目立たなくなっている。つまり、プロミネンスを置くところでは高い部分がより高くなる。その語に続く部分の高さは目立たなくなる。

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

4.3. 日本語の頭高型と中国語の第 4 声との関わり

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

従って、母語のアクセントの干渉で、中国語母語話者は日本語において打ち消しを表す部分「ない」を発音する時、母語の打ち消ししている字「不[pu⁵¹]」のピッチ曲線のように発音してしまう。

4.4. 日本語の平板式と中国語の第 1 声との関わり

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

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

¹ ここでは、アクセントを広義の意味で用いる。

これに対して、中国語の名詞においては、3音節以上の音節を発音する際に日本語の4音節「マユダマ」、5音節の「イキモドリ」、6音節の「マユツバモノ」などのように、下がり目がなく、平坦に発音する名詞は殆ど存在しない。名詞以外でも「开开心心[k^hai⁵⁵k^hai⁵⁵cin⁵⁵cin⁵⁵]、稀稀拉拉[ci⁵⁵ci⁵⁵la⁵⁵la⁵⁵]」のような副詞も少数である。

また、第1声の声調符号が付いている漢字は単独で発音される時、「媽」[ma⁵⁵]のように超高平となるが、いったん二音節のセットになった時、平坦に発音するのではなく、「音節の間に波の山を形成したり、波の上昇部分や下降部分を形成する傾向が強く見られる(平井 2012:102)。

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

4.5. 日本語教材におけるアクセントの解説との関わり

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

(1) 「総合日本語」における動詞打ち消しのアクセントの解説

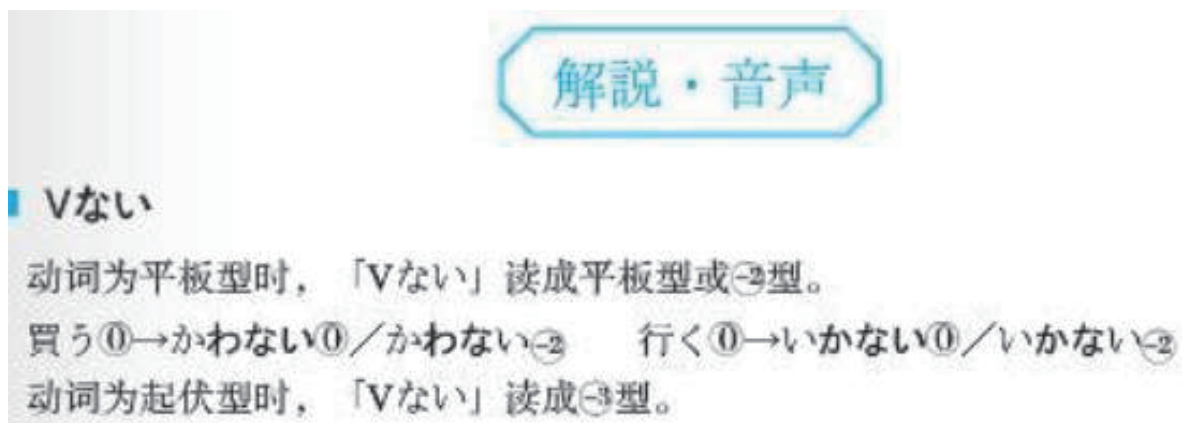


図 3: 「総合日本語」における動詞打ち消しのアクセントの解説

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

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

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

(2) 「新編日本語」における動詞打ち消しのアクセントの解説

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

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

5. まとめ

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

参考文献

- 平井勝利 (2012) 『教師のための中国語音声学』 白帝社
- 彭広陸・守屋三千代など (2009) 『総合日本語第一冊』 北京大学出版社
- 加藤重広・安藤智子 (2016) 『音声学講義』 研究社
- 窪園晴夫 (1998) 『音声学・音韻論』 (西光義弘編「日英語対照による英語学演習シリーズ」1) くろしお出版
- 劉佳琦 (2008) 「東京語の動詞・複合動詞アクセントの生成について」 『2006年清華大学日本言語文化国際フォーラム論文集』 清華大学出版社 pp. 424-439
- 松森晶子・新田哲夫・木部暢子・中井幸比古 (2012) 『日本語アクセント入門』 研究社
- 斎藤純男 (1997) 『日本語音声学入門』 三省堂
- 蔡全勝 (1983) 「中国人に見られる日本語アクセントの傾向」 『在中華人民共和国日本語研修センター紀要 日本語教育研究論纂』 第1号 pp. 26-31
- 周平・陈小芬 (2009) 『新編日语 第一冊』 上海外语教育出版社

自己モニターを伴う日本語アクセントの産出訓練の効果

—中国語母語話者を対象として—

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

1. はじめに

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

2. 先行研究

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

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

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

3. 方法

3.1. 実験協力者

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

3.2. 実験対象語

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

表 1: 実験対象語

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

3.3. 手続き

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

表 2: 実験の流れ

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

3.4. 産出訓練

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

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

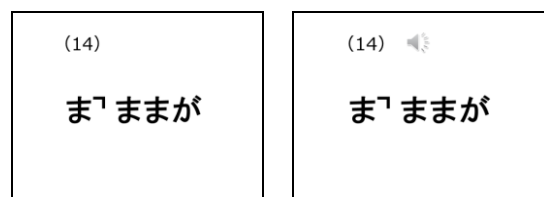


図 1: 産出訓練の呈示画面の例

3.5. テスト

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

4. 結果と考察

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

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

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

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

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

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

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

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

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

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

表 5: 産出訓練グループによる pre-test の誤用数と誤用率(%)

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

「ア型」=「アクセント型」、「—」=「該当なし」

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

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

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

5. おわりに

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

謝辞

本稿は、文部科学省科学研究費：基盤（B）「海外日本語学習者音声アーカイブの構築・分析とWEB韻律学習支援ツール開発」（課題番号：17H02352）による成果の一部である。

参考文献

- 小河原義朗（1998）『外国人日本語学習者の発音学習における自己モニターの研究』（博士論文，東北大学大学院文学研究科）
- 王睿来（2016）「中国語母語話者の日本語名詞アクセントの産出：アクセント核の情報とモデル音声の有無による影響に着目して」『第11回国際日本語教育・日本研究シンポジウム予稿集原稿』.
- 王睿来（2017）「中国語母語話者による日本語複合名詞アクセント産出：学習歴と単純名詞アクセント産出の影響」『ことばの科学研究』18, 31-49.
- 王睿来・林良子・磯村一弘・新井潤（2017）「中国語母語話者による日本語名詞アクセントの習得：知識・産出・知覚の関係から」『中国語話者のための日本語教育研究』7, 61-75.
- 雷寶茵（2007）「香港人広東語母語話者の複合名詞アクセントの生成について」『「日本語教育と音声」研究会第7回：創立125周年記念特別シンポジウム予稿集原稿』.
- 柳悦（2010）『中国人日本語学習者の複合名詞アクセント習得の縦断的研究：知識・発音・知覚の比較を中心に』（博士論文，首都大学東京大学院人文科学研究科）
- 陳冠霖（2015）「台湾人日本語学習者の2モーラアクセントについて：発音、聞き取り、記憶を中心に」『第29回日本音声学会全国大会予稿集』, 24-29.

L2 英語音声の評定における評定者の母語の影響

小西 隆之 (早稲田大学大学院) 近藤 真理子 (早稲田大学)
tkonishi@aoni.waseda.jp, mkondo@waseda.jp

1. 概要

本研究では、第二言語 (L2) 英語音声コーパス (J-AESOP) に収録されている 183 名の日本人英語学習者による *The North Wind and the Sun* の読み上げ音声を計 5 人の英語母語話者および非母語話者が「セグメントの正確さ」などの 4 項目に関して評定したものの結果を比較分析し、異なる母語を持つ評定者間の評定値一致率や、評定が一致しやすい項目を明らかにした。

2. 本研究の目的

第二言語音声習得研究は、習熟度の異なる話者の比較分析を行い、その共通性と習熟段階における変化を検証するのが一般的である。この際、従来研究においては被験者募集のときに海外滞在年数などの明確な指標を用いて初級群と上級群を区分し、その二群の比較分析を行うという手法が用いられてきた。

しかしながら近年の L2 コーパスの普及により、中級群も含めた多様な学習者群をデータとして解析することが可能となった。L2 学習者は言語背景、教育環境や習得適性 (aptitude) 等の個人内要因も影響し、極めて多様な集団を形成する。特に L2 音声習得においてはこの多様性が顕著であると言える。

このような多様な集団を対象とした習熟度別の比較を行うためにはコーパスに評定値を付与する必要がある。Saito et al. (2015) はフランス語母語話者の L2 英語に 7 項目の評定値 (「個々の子音や母音の発音は適切か」、「アクセント、イントネーションは適切か」等) を付与し、音声学者と非音声学者各 10 名の評定値の一致率を検証した結果、全ての項目について高い一致率を得た (クロンバックの $\alpha > 0.9$)。

昨今では英語母語話者よりも L2 英語話者人口の方が圧倒的に多く (Crystal, 2003; Bolton, 2004 等)、英語母語話者が判断する理解容易度 (intelligibility) だけでなく、L2 英語話者、とりわけ異なる言語背景を持つ L2 英語話者の理解容易度も考慮する必要がある。

そこで、本研究は、日本語母語話者の L2 英語音声の評定を対象とし、L2 音声の評定における評定者の母語の影響を検証することにより、L2 音声習得研究における評定基準の指標となることを目的とする。

3. 分析

評定対象としたのは J-AESOP に収録されている様々な習熟度の日本人英語学習者 183 名のデータで、評定対象とした音声は PC スクリーンに呈示された「*The North Wind and the Sun*」を被験者が音読したものである。評定者の個人内誤差を解消することと、特定箇所の

数 0.62 - 0.80)の方が、非母語話者間(すなわち今回の分析では日本語母語話者-スペイン語母語話者間)の相関(相関係数 0.58 - 0.74)よりも高いという傾向が示された。

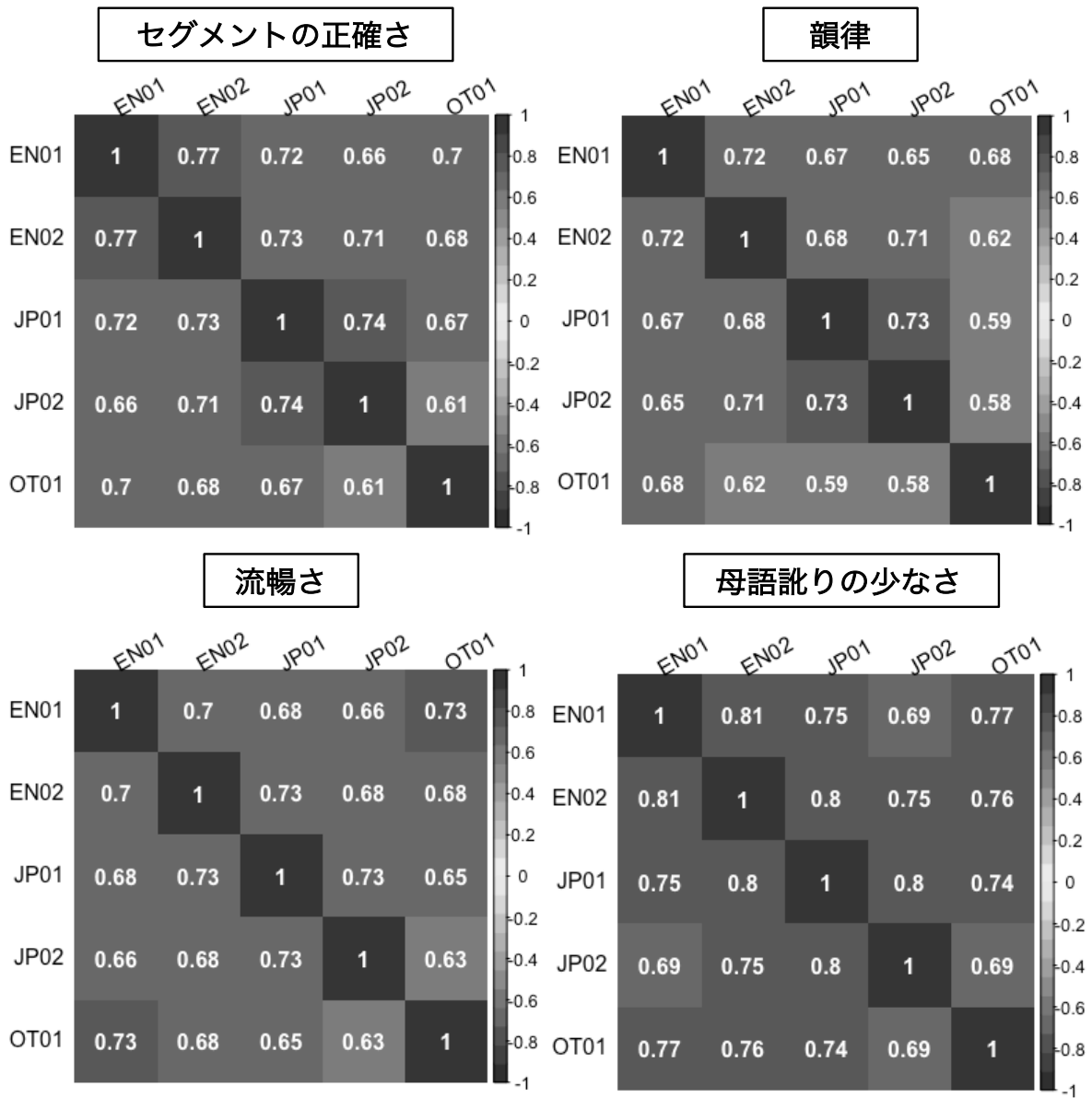


Figure 1 多重相関分析の結果

EN:英語母語話者 JP:日本語母語話者 OT:スペイン語母語話者 01,02:評定者 ID

次に、各項目の評定者間相関を比較してみると、母語訛りに関する評定値の相関係数(0.69 - 0.81)が最も高いことがわかる。母語訛りに関する評定は今回の評定作業において最も直感的に判断される評定項目であることから、音声学を専門とする評定者であっても、音声学的知識を用いた評定よりも直感的な評価の方が一致率が高くなることが示唆される。

また、セグメントに関する評定値一致率(相関係数 0.61 - 0.77)が韻律に関する評定値一致

率(相関係数 0.58 - 0.73)よりも高かったことから、英語のセグメントに関する指標は母語の異なる学習者間においても比較的に共有しており、また、それは英語母語話者のものとも近い一方で、韻律に関する判断は母語の影響を受けやすいという傾向が示された。

5. 今後の展望

今回の分析は既に評定作業を終えた 5 名の評定者のデータを用いて行った比較的の小規模なものであったが、本研究の結果から、L2 英語の評定において英語母語話者間の高い評定値一致率が得られるという傾向は示された。しかしながら、日本語母語話者間の高い一致率は、評定者同士の母語の一致による影響か、または、評定者の母語と評定対象だった L2 英語学習者の母語が一致に起因するものか、どちらの要因が大きいかは今回の分析結果から明らかにすることはできない。今後、他の評定者による評定が終わった時点で再度分析を行い、日本語母語話者以外の非英語母語話者同士(例えばスペイン語母語話者同士)の評定値一致率を日本語母語話者同士の評定値一致率と比較することにより、評定者と異なる母語を持つ L2 英語話者の音声でも、評定者の母語が一致していれば評定値一致率が高くなるかどうかを検証する。

6. 謝辞

本研究は科研費若手研究(B) (No. 17K13513)および基盤研究(B) (No. 15H02729)の助成を受けている。

参考文献

- Bolton, K. (2004). World Englishes. In A. Davies & C. Elder (eds), *Handbook of Applied Linguistics*. Oxford: Blackwell, 367-396.
- Crystal, D. (2003) *English as a Global language*. Cambridge: CUP.
- Saito, K., Trofimovich, P. & Isaacs, T. (2015) “Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels”, *Applied Psycholinguistics* 37(02), 217-240

日本語での中和環境における英語音の知覚 — *Who is Fu?*

青柳 真紀子 (獨協大学) Yue Wang (Simon Fraser University)
aoyagi@dokkyo.ac.jp, yuew@sfu.ca

1. はじめに

本研究は、母語の音韻規則が外国語音の知覚におよぼす影響を探るものである。具体的には、日本語話者が英語音 /f/ と /h/ を同定する際に、日本語の音韻規則 (fu-hu の中和) の影響を受け、/_u/ の環境で難しくなることについて取り上げ、その要因を検討する。

外国語音を聞く際、母語における最も近い音素に知覚同化させることが知られている (Best & Tyler, 2007)。簡単に言えば、外国語の二つの音素が母語でも二つの音素に対応していると聞き分けは容易になり、ひいては産出も習得しやすい (Flege, 1996)。英語の /f/ と /h/ は、それぞれ日本語の /ϕ/ と /h/ に知覚同化され、/f/-/h/ の対立は概して /ϕ/-/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/ → [ϕ] となり、/ϕ/ と /h/ とともに [ϕ] 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* (/fa/-/ha/) の同定は容易である一方、*Fu-who* (/fu/-/hu/) の同定は困難となり得る。

この /fu/-/hu/ の中和が絡む困難についての記述は見られるが (竹林, 1997; Kondo *et al.*, 2015 等)、知覚面での実証的な研究は多くない。そのうち、Lambacher *et al.* (2001) は日本人大学生の /f/ と /h/ の同定率が u の環境で低いことを示したが、その要因についてはほとんど触れていない。このように、母音環境が影響することは分かるが、そもそも日本語話者が摩擦音 /f/-/h/ そのものの違いが判別できるかどうかはよく分かっていない。

本研究では、まず、/f/ と /h/ を含む語の同定に加えて /f/ と /h/ 単独の同定を検証をし、次に、/_u/ の環境 /fu/, /hu/ から抽出した /f/-/h/ の同定能力が、元の語全体 /fu/, /hu/, つまり /u/ の提示でどのように変化するかを検証する。これにより、日本語話者が示す /fu/, /hu/ の同定の困難について、摩擦音の音声処理とそれに対する音韻知識関与の点から考察する。

2. 先行研究

2.1. 無声唇歯摩擦音 /f/ と無声両唇摩擦音 /ϕ/

英語等の /f/ は日本語において /ϕ/ で代用されることが多い。現代日本語では /ϕ/ は /hu/ の異音 [ϕu] としてウ段でのみで存在するが、外来語では /f/ の代用子音としてすべての母音とともに使われている¹。このように、/f/ は /ϕ/ で代用され、英語の /f/-/h/ の対立は概して保たれる (例: ファ vs ハ)。しかし、既述のように /ϕ/ は元来 /_u/ の環境での /h/ の異音であることから、/hu/ と /ϕu/ はともに [ϕu] として中和し、対立を失う (フ vs フ)。唇での摩擦の有無や程度は対立を作らないので、[h] と [ϕ] は相互交換的で音声的には両者の間のどのあたりでも出現する。このように、日本語において調音的にも聴覚的にも区別す

¹ /f/ を u の母音つき /ϕu/ で代用する場合もあるが (例: fan > ϕu.an ファン, film > ϕu.i.ru.mu フィルム, felt > ϕu.e.ru.to フェルト), 今日, u が二重母音化, さらには脱落して, /ϕ/ は単独で /f/ の代用子音として機能することが多い (例: fan > ϕan ファン, film > ϕi.ru.mu フィルム, felt > ϕe.ru.to フェルト)。

る必要がないことから、これが転移して英語においても /fu/ と /hu/ の同定は難しくなり、これは /fa/ と /ha/ など他の母音の場合の同定と比べても明らかである。

2.2. 無声唇歯摩擦音 /f/ と無声喉頭摩擦音 /h/

英語の /f/ は、音響的にはスペクトルピークや傾きを含むスペクトル情報、振幅、CV 振幅比、持続時間、CV 推移などの点から特徴づけられている (Jongman *et al.*, 2000) 。しかし、各種摩擦音間の音響的比較および知覚的比較の研究にはほとんど /h/ が含まれない (McMurray *et al.*, 2011 等) 。 /h/ は後続母音の無声開始部のようなもので特定の狭窄点をもたないことから、摩擦音としてのある一定の音響的特長に欠けるためであろう (Ladefoged, 2005) 。このように音響的にも知覚的にも /f/ と /h/ が直接比較されることは多くなく、ましてや /fu/ と /hu/ 限定の環境ではほとんど見られない。

日本語話者の英語発音の注意点として /f/ はよく上げられ、両唇 [ϕ] ではなく唇歯 [f] の摩擦であることが指導される (牧野, 2005 等) 。その際 /h/ との混同が触れられることもあるが、/_u/ の環境 /fu/ - /hu/ での違いへの注意喚起や、その知覚についての言及は多くない。

調音面において、Kondo *et al.* (2015) は、日本語話者にとって英語の /f/ - /h/ は /fu/ - /hu/ の音素配列に特化して困難であるとしている (Table 2(b)) 。実際、96 人による読上げ発話において /f/ と /h/ の混同がよく見られることが示されているが、混同の母音環境については特に触れられていない。Nogita (2010) も日本人英語学習者 4 人の英語読上げ実験で *food-who'd* における /f/ と /h/ の混同を示している。

本題の知覚面では、Lambacher *et al.* (2001) が、日本人大学生 100 余名に対する英語子音聞き取り実験により、/f/ と /h/ は、u の環境 /_u/, /u_u/, /u_/ で同定正答率と識別感受性 (*d'* スコア) が低いことを示しているが、その要因については分析されていない。

3. 実験

3.1. 目的

日本語での中和環境 (/ϕu/ - /hu/ = フ vs フ) に対応する英語音の対立 /fu/ - /hu/ の知覚について、日本語話者と英語話者を対象に調査をする。まず、/f/ と /h/ を含む語 (実験 1)、および語から抽出した摩擦音部分のみ (実験 2) について語彙、子音の同定実験を行う。次に、語から抽出した母音部分のみについて、先行子音の検知実験を行う (実験 3)。

実験 1 で /_u/ の環境の影響を確認した後、実験 2 で摩擦音単独での同定、さらに、実験 3 で後続母音部に残る情報による先行子音の検知について検証する。焦点は、/fu/ と /hu/ から抽出した /f/ と /h/ の摩擦音の同定が、元の語全体 /fu/ と /hu/ の提示、つまり /_u/ の環境の提示によってどのように変化するかを見ることである。これにより、母音環境を限定した中和という音韻規則が、子音そのものの知覚に与える影響を考察する。

3.2. 方法

<使用した語> 実験には、表 1 に示すように /f/ と /h/ から始まり、/_i/, /_u/, /_a/, および問題の母音 /_u/ との境界を示すための /_oo/ を含めた 4 母音が後続する 1 音節対語、さらに、渡り音をはさんで /u/ が続く (/_ju/) 1 音節対語の計 5 対 10 語を使用した。また、*see - she, sip - ship, said - shed, sue - shoe, sock - shock* をダミーとして使用した。

表 1. 使用した語²

	/i/	/u/	/a/	/ou/	/ju/
/f/	<i>fee</i>	<i>Fu</i>	<i>fah</i>	<i>foe</i>	<i>few</i>
/h/	<i>he</i>	<i>Who</i>	<i>hah</i>	<i>hoe</i>	<i>Hugh</i>

<録音> 成人北米英語母語話者 4 人 (男 2, 女 2) が実験語を単独で読み上げたものを録音した (マイク Shure KSM, サンプリング周波数 44.1kHz, 量子化 16bit).

<刺激音> 以下の 3 種類の刺激音を作成した.

実験 1: 語全体 表 1 の 10 語. 語全体を 65dB に標準化した.

(1) *fee-he*, *Fu-who*, *fah-hah*, *foe-hoe*, *few-Hugh*



実験 2: 摩擦音部 (a) 上記 *fee-he*, *fah-hah*, *Fu-who* から摩擦音部 [f] と [h] を抽出したもの, および (b) [f] と [h] を単独で発したもの, 計 8 刺激. (a) は波形上で摩擦開始部分から後続母音の最初の有声周期開始部分までとした. 摩擦部全体を 45dB に標準化した.

(2a) *f(ee)*, *h(e)*, *f(ah)*, *h(ah)*, *f(u)*, *wh(o)*

(カッコ内は分離された母音)

(2b) *f(iso)*, *h(iso)*

(iso: 単独での発話)

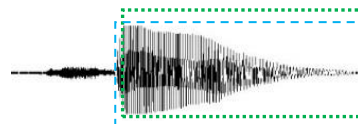


実験 3: 後続母音部 実験語のうち *fee-he*, *fah-hah*, *Fu-who* から上記(2)の摩擦音部分を除いた母音部分で波形上, (a) 母音の 1 つ目の周期への立ち上がりから母音終了点までと, (b) 3 つ目の周期への立ち上がりから母音終了点までを抽出したもの, 計 12 刺激. (a) の方が, 先行子音 (f または h) の同時調音的情報をより多く含んでいると考えられる.

(3a) *(f)ee_1stV*, *(f)ah_1stV*, *(f)ah_1stV*, *(h)ee_1stV*, *(h)ah_1stV*, *(h)ah_1stV*

(3b) *(f)ee_3rdV*, *(f)ah_3rdV*, *(f)ah_3rdV*, *(h)ee_3rdV*, *(h)ah_3rdV*, *(h)ah_3rdV*

(カッコ内は分離された子音)



<手順> 作成された刺激は以下のとおり実験ソフト *Paradigm* 上に準備された.

実験 1 語全体: 計 160 刺激 ([10 語 x 4 話者] x 2 回 + ダミー 80 刺激 = 160)

実験 2 摩擦部のみ: 計 96 刺激 ([2 子音 x 4 母音環境 x 4 話者] x 3 回 = 96)

実験 3 母音部のみ: 計 96 刺激 ([3 母音 x 2 子音環境 x gating2 段階 x 4 話者] x 2 回 = 96)

刺激は被験者毎にランダムに提示され, 直後に選択肢 (例: *fee* *he*) が画面に表示される. 被験者はヘッドセットで刺激を聞き, 画面上で強制二者択一の同定判断をする.

被験者は日本語母語話者大学生 14 人 (男 6, 女 8) と北米英語母語話者大学生 14 人 (男 2, 女 12) で, 手順説明の際に実験語の意味と音, 特に /f/ と /h/ の対語群 (*who* は綴りは 'wh' でも /h/ で始まる語) であることを確認した. 練習と休憩を入れて合計 45-50 分であった.

² *Fu* は中国人名 (*Mr. Fu*), *fah* は音符名 (ドレミ「ファ」), *hah* は間投詞, *Hugh* は人名.

4. 結果と考察

4.1. 実験1 (fとhを含む語全体の同定)

まず, 英語母語話者 (EngNS) の f と h の対語の正答率は母音環境にかかわらずほぼ 100% であった (*hah* は 99%). これに対し, 日本語話者 (Jpn) は *Fu/who* と *few/Hugh* の正答率が低い (図 1). *Fu/who* は予想通りだが, *few/Hugh* は日本語でも対立があるので (フュー/ヒュー) 予想外であった. 語彙親密度の影響だと思われる反応が見られ, *Fu/who* では聞き慣れない *Fu* に対し頻出の *who* (/h/) の回答率と正答率が高く, 逆に *few/Hugh* では不慣れな *Hugh* に対し頻出の *few* (/f/) が高い. 対語間の反応の偏りが関与することから³, 信号検出論に基づく識別感度 (d') を見ると (図 2), /_u/ と /_ju/ で同定が困難であることが分かる. Jpn 上級者 (TOEIC850 以上) の *few/Hugh* は比較的よいことから, *Fu/who* の困難さが確認される.

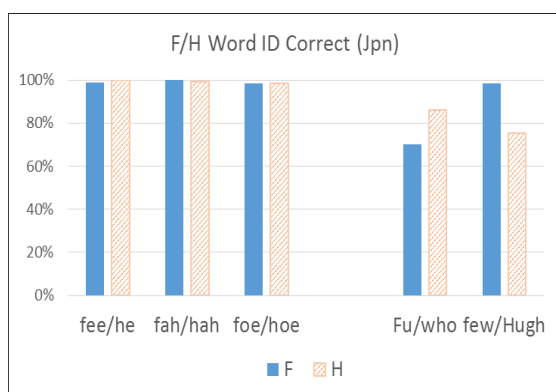


図 1. f/h 語彙同定の正答率 (Jpn)

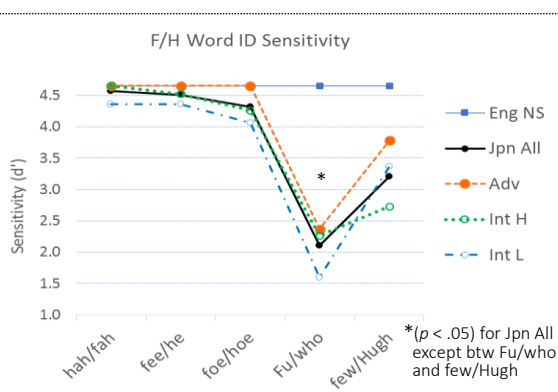


図 2. f/h 語彙同定の識別感度 (EngNS, Jpn)

4.2. 実験2 (抽出されたfとhの同定)

摩擦音部分のみでも EngNS の正答率と識別感度は母音環境に関係なくほぼ完璧である. Jpn は, *Fu/who* から抽出された *f(u)-wh(o)*, および単独発話された *f(iso)-h(iso)* で感度が低下するが, 全般として摩擦音のみでもかなり子音を同定できることがわかる (図 3, 4). 本研究の焦点である語全体と抽出子音単体での感度の差については, 別に 4.4 節に示す.

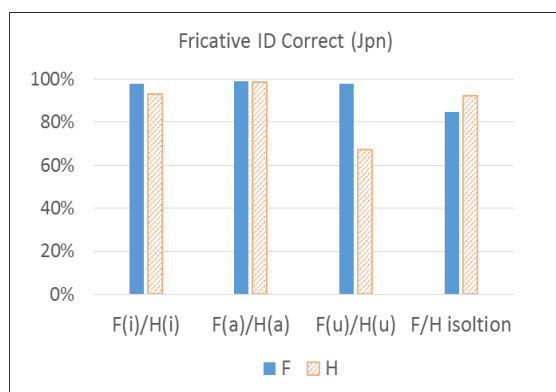


図 3. 抽出された f/h の同定正答率 (Jpn)

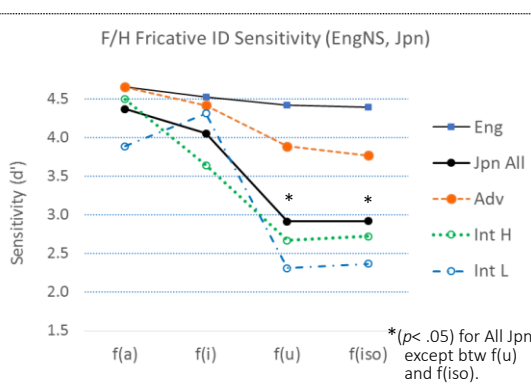


図 4. 抽出された f/h の識別感度 (EngNS, Jpn)

3 仮に常に *f* と回答すると, *f* の正答率は 100% になるが *h* は 0% となる. この *f* の 100% は *f/h* の完璧な判別力を意味しない. 一方, *f* の正答率が 100% で *h* も 100% であれば, *f* の 100% の意味合いは先の例とは全く異なる. よって, *f* と *h* 単体での正答率ではなく, 両者の正答と誤答の兼合いが識別感度には重要となる.

4.3. 実験3（抽出された母音部によるfとhの検知）

母音部から先行子音 f/h 検知する課題の正答率は概して低くなるが，EngNS では全母音で，Jpn でも(f)u の場合を除いて偶発率以上となる．識別感度を見ると，EngNS は母音にかかわらず，母音が進んで先行子音の情報が減少するにつれ子音検知が難しくなる自然なパターン見られるが（図5），Jpn は(f/h)u ではこれが少なく，(f/h)eを除いて入渡り部の音響情報への感度が鈍い．特にここでも/_u/の環境 (f/h)u での感度の悪さが見られる（図6）．

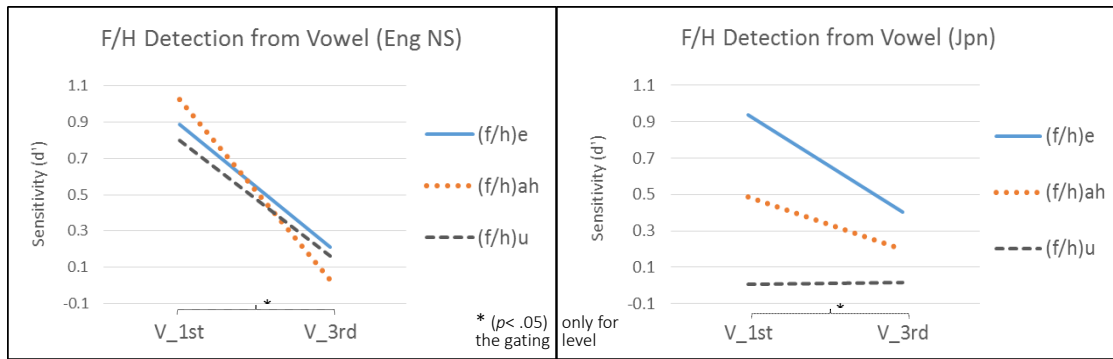


図5. 母音から先行 f/h を検知 (EngNS)

図6. 母音から先行 f/h を検知 (Jpn)

4.4. 子音のみ (C) と子音-母音 (CV) の知覚

通常，子音の同定は子音のみ (C₁) よりも後続母音を伴った (C₁V) 方が容易である (C₁ < C₁V) (C₁V と C₁の C₁は抽出前後の同じ子音)．例えば，[f]単独よりも[fa]の方が推移部のキューなど音響的情報が多く，また音節語として成立し，経験と予測からくる処理能力の向上も期待される (McMurray *et al.*, 2015)．ところが，Jpn は，/_u/の環境では子音のみの方が識別感度が高くなる (C > CV) (表2 下段)．つまり，*fah/hah* や *fee/he* から抽出された[f]，[h]の識別感度は，元の語全体が提示されると上昇する一方で，*Fu/who* から抽出された[f]，[h]では，単体ではできていた識別が，それが元々含まれていた抽出元の語全体が提示されるとできなくなってしまうと言える．一部のみでできていた識別が，それを含む全体の提示によってできなくなるということは，一般的な予測に反する．これは，音声的に[f]と[h]がある程度識別できていても，/_u/の提示により/f/と/h/は異音になるこ

表2. “抽出されたCの同定”と“抽出元の語全体の同定”の感度比較 (Jpn)

	C		CV	
	f/h		fah/hah	
Jpn All	4.37	↗	4.57	
Advanced	4.65	↗	4.65	
Intermed H	4.50	↗	4.65	
Intermed L	3.89	↗	4.36	

	C		CV	
	f/h		fee/he	
Jpn All	4.05	↗	4.51	
Advanced	4.42	↗	4.65	
Intermed H	3.64	↗	4.52	
Intermed L	4.31	↗	4.36	

	C		CV	
	f/h		Fu/who	
Jpn All	2.91	↘	2.10	
Advanced	3.89	↘	2.37	
Intermed H	2.67	↘	2.26	
Intermed L	2.31	↘	1.60	

↗ : CよりCVで感度上昇
↘ : CよりCVで感度低下

とから、異音識別の鈍化が引き起こされている可能性が考えられる。また、これは Boomershine *et al.* (2008) の示した音素と異音の知覚感度の違いと類似するものである。

5. 結論と今後

これまで見てきたように、語全体でも、摩擦音部のみでも、母音部のみでも、日本人被験者は他の母音環境と比べて /_u/ の環境で /f/-/h/ の識別感度が低い (それぞれ 4.1, 4.2, 4.3)。これに加え、/_u/ の環境で、同じ信号 C₁ の同定感度がその抽出元全体 C₁V で鈍化すること (C₁ > C₁V, つまり $f(u)-h(u) > Fu-who$) が見られたことは (4.4)、/f/-/h/ について、より基礎的な音声処理が、より高次の音韻規則の知識によって阻害されていることを示すと考えられる。また、通常 C₁V は C₁ の知覚を容易にするが、逆に阻害する一例でもある。

別の見方として、日本語話者は摩擦音 /f/-/h/ のみの識別時には音声として聞いていない可能性がある。Werker *et al.* (1985) が、英語話者がセイリッシュ語のある種の子音を言語音声としては同定できないが非言語音声としては同定できることから別モードでの知覚を示唆していることに類似するが、本研究のデザインではそれを検証するに至らない。

また、/_u/ の環境の提示が /f/-/h/ の異音化と識別鈍化を引き起こすという仮説を検証するには、弁別精度を見る実験、子音部分に母音情報を加算していく実験、また、/_u/ 環境の擬似提示 (音響ではなく「期待」の提示) によるプライム効果を確認する実験等、別の手法が必要となろう。現在、/f/-/h/ の音響差異の母音環境による変化を分析中だが、この結果とそれに基づいた知覚実験も音響と知覚の関係を探る基礎項目として必須である。

謝辞：本研究の一部は Simon Fraser University, Language and Brain Lab と Department of Linguistics の協力と助成を受けて行われ、また、河合裕美先生にも実験実施にご協力いただきました。ここに感謝します。

参考文献

- 竹林滋 (1997) 『英語音声学』東京：研究社。
- 牧野武彦 (2005) 『日本人のための英語音声学レッスン』東京：大修館。
- Best, C., & Tyler, M. (2007) "Nonnative and second-language speech perception," In O.-S. Bohn & M. Munro (eds.) *Language Experience in Second Language Speech Learning*. Amsterdam: John Benjamins.
- Boomershine, A., Hall, K., Hume, E. & Johnson, K. (2008) "The impact of allophony vs contrast on speech perception," In B. Avery, E. Dresher & J. Rice (eds) *Phonological Contrast*. N.Y.: Mouton de Gruyter.
- Flege, J. E. (1996) "Second language speech learning: Theory, findings, and problems," In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-linguistic Research*. Timonium: York Press.
- Jongman, A., Wayland, R. & Wong, Serena. (2000) "Acoustic characteristics of English fricatives," *J. of Acoustical Society of America* 108 (3), Pt. 1, 1252-1263.
- Kondo, M., Tsubaki, H. & Sagisaka, Y. (2015) "Segmental variation of Japanese speakers' English: Analysis of "the North Wind and the Sun" in AESOP Corpus," *J. of Phonetic Society of Japan* 19 (1), 3-17.
- Lambacher, S., Martens, W., Nelson, B. & Berman, J. (2001) "Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias," *Acoustical Science and Technology* 22 (5), 334-343.
- McMurray, B. & Jongman, A. (2011) "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," *Psychological Review* 118 (2), 219-246.
- McMurray, B. & Jongman, A. (2015) "What comes after /f/? Prediction in speech derives from Data-Explanatory Processes," *Psychological Science OnlineFirst*, published on Nov. 18, 2015.
- Nogita, A. (2010) "Do Japanese ESL learners' pronunciation errors come from inability to articulate or misconceptions about target sounds?" *Working Papers of the Ling. Circle of U. of Victoria* 20 (1), 82-116.
- Werker, J. & Logan, J. (1985) "Cross-language evidence for three factors in speech perception," *Perception & Psychophysics* 37 (1), 35-44.

韓国語母語話者による英語の母音の知覚判断

-後続子音の影響について-

韓 喜善 (大阪大学) , 野澤 健 (立命館大学)
kenkyuhhs@gmail.com, t-nozawa@ec.ritsumeai.ac.jp

1. 本研究の背景と目的

これまで、英語の母音に関する知覚実験の多くは、子音の影響を最小限にするために、声門摩擦音と閉鎖音の間に母音を入れた音環境を中心に検討されてきた。また、非母語話者を対象とした場合、母音に後続する子音の影響について調査した論考は少ない。本研究では、韓国語母語話者による英語の /i, ɪ, ε, æ, α, ʌ/ の 6 つの母音の知覚判断について調査した。その際、閉鎖音 /t/、鼻音 /n/、流音 /l/ を後続子音として配置し、後続子音の調音様式による母音の知覚判断の違いについて検討した。

2. 先行研究

2.1. 韓国語と英語の短母音

母音の数や種類に関しては、研究者によって見解が異なる場合もあるが、現代の韓国語の短母音は /i, e, a, o, u, ʌ, u/ の 7 つという見解 (Oh 1996, Kang 2013, Hwang 2015 等) が主流である。米英語の短母音については、Ladefoged (2003, 2006) では、/i, e, ɪ, ε, æ, α, ʌ, ʊ, u, o, ə, ə/ の 12 個という見解が示されている。両言語の母音は、母音の音素の数が異なるだけでなく、互いの母音の舌の前後や舌の高さの配置が異なっている (진 2005)。

2.2. 韓国語母語話者による英語の母音の生成と知覚

英語の /i, ɪ, æ, ε/ の母音はすべて前舌母音ではあるが、前舌母音 (F2) としての度合いに差があり、口の開き (F1) の度合いも異なる (Ingram and Park 1997, Ladefoged 2006 等)。そのため、これらの子音を聞き分けるためには、それぞれの F1 と F2 の周波数に注目する必要がある。韓国語母語話者の英語の母音の知覚判断については、/i/ と /ɪ/ との間、/æ/ と /ε/ との間、の弁別が困難であることが報告されており (Ingram and Park 1997, Flege et al. 1997, Frieda and Nozawa 2007 など)、/i/ と /ɪ/ を韓国語の /i/ として /æ/ と /ε/ を韓国語の /e/ に知覚しているという (Frieda and Nozawa 2007)。生成においても、Ingram and Park (1997) によると、韓国語母語話者の /i/ と /ɪ/ の F1 と F2 は同等であり、/æ/ と /ε/ の F1 と F2 も同等であると報告している。これらの母音の判断には、母音長に注目すれば、母音の判断率が高くなるのが日本語母語話者の例から知られているが (Ingram and Park 1997, Frieda and Nozawa 2007)、韓国語母語話者は学習歴や現地滞在歴に関わらず、このようなストラテジーを採用しないため、英語の母音の知覚と生成の問題は継続する傾向があるという。

一方、後舌の /α/ と中舌寄りの /ʌ/ については、韓国語母語話者がこれらの母音の弁別がほぼ可能であることも報告されている (Frieda and Nozawa 2007)。

このように、韓国語母語話者による英語の /i, ɪ, æ, ε/ の音声の難易度に関しては、先行研究間で一貫した結果を得ているが、いずれも後続母音が閉鎖音で終わる環境での報告に限られている。

3. 本研究の課題

本研究では、先行研究で取り上げた英語の6つの母音 (/i, ɪ, æ, ε, ɑ, ʌ/) について、後続する子音の影響を調査する。後続子音として閉鎖音 (/t/)、鼻音 (/n/)、流音 (/l/) を選び、それぞれの環境を有する英語の母音の弁別について比較した。先行研究で取り上げられていた語末閉鎖音の他に鼻音と流音を調査項目として加えたのは、母音に後続する鼻音のフォルマントが母音の F1 付近に現れることにより、母音の高さの知覚が影響されるものと予想されるためである。特に、/æ/ は鼻音の前では上方 (F1 の周波数の低下)、前方 (F2 の周波数の上昇) に移動することが知られているが (Labov 2001)、このことが韓国語母語話者の英語の知覚に影響することも予想される。また /l/ が後続すると F2 が継続して降下するため、他の子音環境とは異なる響きが生じることが予想される。これらの仮説について調査を行い、その検証を行う。

4. 実験の手順

4.1. テスト語

米英語の母語話者4名 (20代の女性、大学生) による音声を使用した。4名の話者の出身地は、ニューヨーク州2名、カリフォルニア州1名、ウィスコンシン州1名である。オーバーン大学 (Auburn University、米国アラバマ州) の実験室内のコンピュータのHDに直接録音した。/hVt/、/pVt/、/pVn/、/pVl/ の環境で発した24語を刺激音の素材として使用した。テスト語の中には、人名 (Pete、Pell) や無意味語 (het、pon、pul) も含まれる。

表1 テスト語のリスト

	i	ɪ	ε	æ	ɑ	ʌ
hVt	heat	hit	het	hat	hot	hut
pVt	Pete	pit	pet	pat	pot	putt
pVn	peen	pin	pen	pan	pon	pun
pVl	peel	pill	Pell	pal	pol	pul

4.2. 音響分析

収集した音声は、母音長、母音の中間時点での F1 と F2 を測定した¹⁾。長さに関しては、/i/ は /ɪ/ に対して 1.1~1.3 倍長く、/æ/ は /ε/ に対して 1.1~1.7 倍長く、/ɑ/ は /ʌ/ に対して 1.1~1.6 倍長かった。特に /pVn/ の環境が母音種による長さの割合の差が他より大きかった。

¹⁾ /pVl/ の環境に関しては母音と /l/ との分節ができないため、長さに関しては母音と /l/ の区間を測定し、フォルマントに関しては母音と /l/ 区間の 1/4、2/4、3/4 の時点についてそれぞれ F1 と F2 の測定を行った。

フォルマントに関しては、/hVt/と/pVt/とでは、それぞれの母音の F1 と F2 がほぼ同様の周波数を示し、語頭子音の調音位置や調音様式による影響はなかった。

/pVn/では、/hVt/と/pVt/に比べて、/i, ɪ, ε, æ/は F2 が高く、/ɑ/と/ʌ/は F2 が低かった。特に、/æ/の F2 (2407Hz) は/hVt/ (1789 Hz) と/pVt/ (1897 Hz) の環境より著しく高く、Labov (2001) の結果と一致する。F1 に関しては、/i, ɪ, ε, ʌ/は/hVt/と/pVt/の環境の F1 とほぼ同じかそれに近い周波数を示しているが、/æ, ɑ/では/pVn/の環境においては F1 の周波数が低くなっている。/pVn/は、/pVt/環境と調音場所が一致するため、調音様式（口音か鼻音か）による影響と推測できるが、今後さらに詳しい検討が必要である。

/pVl/の母音と/l/区間の 4/1 時点については、すべての母音の F2 の周波数が低いことから、後寄りに発音されていることがわかる。これは英語の語末に位置する/l/の音声がいわゆる“dark l（軟口蓋化した有声歯茎側面接近音：[ɫ]、軟口蓋化有声歯茎側面接近音：[ɫ̥]）”として発音されることによって、後舌が軟口蓋に接近するという調音が先行する母音の調音段階から始まっているものと解釈できる。舌の高さについては、母音によって傾向が異なり、/i, ɪ/は F1 の周波数が多少高くなり広母音化している。それに対して、/æ, ɑ, ʌ/は F1 の周波数が低くなり、狭母音化している。/ε/は/hVt/、/pVt/の場合とほとんど変わらない。このように、/l/は先行する母音を後舌化すると同時に、狭母音と広母音の差を不明瞭（結果として中舌化）なものにする働きをしていることがわかる。この傾向は子音部への移行が進むほどより顕著になっていく。

4.3. 実験方法

実験は、/hVt/、/pVt/、/pVn/、/pVl/のそれぞれ音環境ごとに行った。実験参加者に、音声刺激を聞き、聞いた母音を選択肢の中から選択する方式で回答させた。選択肢は「4.1.」で説明した通り（表 1 を参照）、/i, ɪ, ε, æ, ɑ, ʌ/の 6 つのいずれかの母音を含む有意義ないし無意味語である。提示した語は音声記号ではなく、英語の綴りで表記したため、音と綴りとの関係を予め説明した。例えば、het は pen と同じ母音であること、putt は put ではなく but と同じ母音を含む語であることを説明した。

刺激音は 2 回試行し、音環境ごとに 48 個（6 母音×4 話者×2 回=48 個）の刺激音に対する評価となる。刺激音はランダムに配列し、子音環境の提示順による影響を考慮し、参加者により、子音環境の提示順番を変えた。4 つの子音環境で計 192 個の刺激音の判断を 30 程度所用した。途中で休憩を 10 分ほど入れた。

4.4. 実験参加者

実験参加者は韓国母語話者 22 名（18-35 歳、男性:6 名、女性:16 名）である。すべての学習者は韓国語を母語とし、近畿圏の大学に在籍している学部生と大学院生である。日本滞在歴は 1 年 3 ヶ月～15 年 3 ヶ月である。英語能力に関しては 1～9 段階（1: very poor、9: very good）の中から選ぶように指示した。

5. 結果と分析

6 母音の各子音環境での平均正答率を表 2~5 に提示する。4 つの子音環境と 6 母音を被験者内要因にした反復測定分散分析の結果、子音環境の主効果 ($F(3, 63)=41.782, p<.001$)、母音の主効果 ($F(5, 105)=35.469, p<.001$)、子音環境と母音の交互作用 ($F(15, 315)=19.140, p<.001$)といずれも有意であった。多重比較の結果、/i/を除くすべての母音が子音環境の影響を受けていることがわかった。/i/では、/pVn/は他の音環境より有意に正答率が低く ($p<.001$)、/hVt/は/pVt/より正答率が低い ($p<.01$)。/ε/では、/hVt/は他の音環境より有意に正答率が低い ($p<.01$)。/æ/では、/hVt/は他の音環境より有意に正答率が高く ($p<.01$)、/pVI/は/pVt/より有意に低い ($p<.01$)。/a/では、/pVn/と/pVI/は/hVt/と/pVt/に比べて有意に正答率が低く ($p<.001$)、/pVn/より/pVI/のほうが正答率が低い ($p<.01$)。/ʌ/では、/pVI/は他の音環境に比べて有意に正答率が低かった ($p<.001$)。以下、/i/と/i/、/ε/と/æ/、/a/と/ʌ/の比較を中心に論を進めていく。

5.1. /i/と/i/

/i/と/i/の音韻の混乱が見られ、先行研究と一致する結果を得た (Ingram and Park 1997、Flege et al. 1997、Frieda and Nozawa 2007 など)。全体的に/i/を含む語の heat (55%)、Pete (47%)、peen (42%)、peel (47%) に対する正答率より/i/を含む語の hit (67%)、pit (84%)、pin (39%)、pill (75%) に対する正答率が高い。実験後に、実験参加者に heat と hit、Pete と pit、peen と pin、peel と pill の区別をどのようにしたかについてインタビューをしたところ、「母音の長さ」という回答が多かったが、実際には母音長に差があっても (4.2 参照)、母音長の影響とは言えない結果となっている。特に、/pVn/では母音長の相対的な差が他の環境より顕著であることを「4.2」で報告した。このように、/i/を/i/に間違えやすいという結果に偏ったのには、綴り字の影響を考慮する必要がある。つまり、例えば heat か hit か判断に迷った際に綴りがより簡単な hit を優先的に選んだ可能性がある。/pVn/の環境においては、/i/を/ε/と間違える場合があった (24%)。/i/と/ε/の F1 の差は、/hVt/と/pVt/に比べて/pVn/の環境で小さい。

5.2. /ε/と/æ/

/æ/と/ε/についても、これら 2 つの音韻を間違えやすいという結果となり、先行研究と一致した (Ingram and Park 1997、Flege et al. 1997、Frieda and Nozawa 2007 など)。音環境によって以下の 2 つの傾向が見られた。/ε/の正答率は、/hVt/ (48%) が/pVt/ (63%)、/pVn/ (69%)、/pVI/ (77%) よりむしろ低い。しかし、/æ/の正答率はその反対で/hVt/ (69%) で最も高い。これは、有意味語 (het) か無意味語 (hat) による影響を考慮する必要があり、/ε/か/æ/かの判断に迷った時に有意味語を優先して選んだ可能性がある。したがって、以降では音環境の影響については/pVt/、/pVn/、/pVI/の間の比較に絞って検討する。

/æ/の正答率は、/pVt/ (47%) より/pVn/ (39%) と/pVI/で低い (32%)。/pVn/で/æ/を/ε/と間違えやすかった理由は、/æ/は鼻音の前では上方 (F1 の周波数の低下) と前方 (F2 の周波数の上昇) に移動し、/ε/に近づいたためと解釈できる。一方、/pVI/でも/æ/は/ε/に回答が集中しているが、ここでは/ε/の F2 が/æ/に近づいているためと解釈できる。しかし、/ε/の正答率に関してはその正答率が高く (78%)、不明な点が残る。

5.3. /ɑ/と/ʌ/

音環境によって2つの傾向に分かれた。まず、/hVt/、/pVt/の閉鎖子音の環境では、72%以上の割合で/ɑ/と/ʌ/の弁別ができています。これは、先行研究と一致する結果である（Frieda and Nozawa 2007 など）。一方、/pVn/では/ɑ/（無意味語 pon）を/ʌ/（pun : 45%）と判断する場合があります。/pVn/では、/ɑ/と/ʌ/がかなり接近していた。この現象は/pVI/の環境においても見られた。また、/pVI/では/ɑ/（poll）を/æ/（pal : 49%）か/ʌ/（無意味語 pul : 23%）に判断するという結果であった。これは、この環境において/ɑ/と/æ/は舌の高さ（F1）で類似し、/ɑ/と/ʌ/は舌の前後（F2）の位置が近い。その結果、pul（/ʌ/）に関しては/ɑ/（47%）と判断しやすかったと見られる。

表2 /hVt/の知覚判断の結果(%)

		選択肢					
		/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
刺激音	heat /i/	55	44	2	0	0	0
	hit /ɪ/	28	67	2	2	0	1
	het /ɛ/	1	7	48	39	1	5
	hat /æ/	2	1	27	69	1	1
	hot /ɑ/	1	0	1	9	77	13
	hut /ʌ/	0	0	3	5	10	82

表3 /pVt/の知覚判断の結果(%)

		選択肢					
		/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
刺激音	Pete /i/	47	52	1	0	0	1
	pit /ɪ/	14	84	1	1	0	0
	pet /ɛ/	11	6	63	14	1	6
	pat /æ/	2	2	48	47	1	1
	pot /ɑ/	0	0	0	13	72	15
	putt /ʌ/	0	1	2	1	10	86

表4 /pVn/の知覚判断の結果(%)

		選択肢					
		/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
刺激音	peen /i/	42	57	1	0	1	0
	pin /ɪ/	21	39	24	4	0	0
	pen /ɛ/	1	14	69	15	1	1
	pan /æ/	2	2	57	39	0	0
	pon /ɑ/	0	1	0	10	44	45
	pun /ʌ/	0	0	1	5	15	66

表 5 /pVl/の知覚判断の結果(%)

		選択肢					
		/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
刺激音	peel /i/	47	51	1	0	0	1
	pill /ɪ/	16	75	5	1	1	2
	Pell /ɛ/	3	8	77	8	1	3
	pal /æ/	2	0	63	32	2	1
	pol /ɑ/	0	0	1	49	27	23
	pul /ʌ/	0	1	2	10	47	41

6. 結論

今回の実験結果は、韓国語母語話者は/i/と/ɪ/、/æ/と/ɛ/の判断が困難であるという点で先行研究と一致する。しかし、/ɑ/と/ʌ/に関してはこれら2つの音韻の弁別が可能である(Frieda and Nozawa 2007)ということが知られてきたが、鼻音が後続する環境や流音が後続する環境ではこれらの音素の区別は容易ではないことが明らかになり、音環境によって音素の区別の難易度は変化することがわかった。また、本研究に参加した実験参加者は全員第2言語として日本語を習得しており、その大半が母音の長短が習得できていることが期待されたが、実際には英語の知覚判断においては母音の長さに注目していないことがわかった。

参考文献

- Ingramm, J. CL and S. Park (1997) "Cross-language vowel perception and production by Japanese and Korean learners of English", *Journal of Phonetics*25:3, pp. 343-370.
- Kang, Y. (2013) "A Corpus-Based Study of Positional Variation in Seoul Korean Vowels," *Japanese/Korean Linguistics* 23.
- Ladefoged, P. (2006) *A Course in Phonetics*, Wadsworth.
- Hwang, H. (2015) "The role of perceived similarity and contrast: English loanwords into Korean and Japanese," *Proceeding of International Conference on Phonetics and Phonology 2015*.
- Frieda, E. and T. Nozawa (2007) "You are what you eat phonetically: The effect of linguistic experience on the perception of foreign vowels," *Language Experience in Second Language Speech Learning in honor of James Emil Flege*, pp. 189-200, John Benjamins.
- Labov, W. (2001) *Principles of Linguistic Change*, Blackwell.
- Ladefoged, P. (2003) 「アメリカ英語」『国際音声記号ガイドブック』, pp. 58-62.
- 이호영(1996) 국어음성학, 태학사.
- 전상범(2005) 전면 개정판 영어음성학개론, 을유문화사.

本研究は科研費基盤研究 C (16K02650、研究代表者：野澤健) の助成を受けたものである。