C1

A preliminary study of the vowel length contrast in Drenjongke

Seunghun J. Lee (International Christian University, University of Venda)
Céleste Guillemot (Daito Bunka University, International Christian University)
Audrey H. Lai (International Christian University)
Honoka Asai (International Christian University)
Kotone Sato (International Christian University)
seunghun@icu.ac.jp, celeste.guillemot@gmail.com,
c201727e@icu.ac.jp, c191048r@icu.ac.jp, c211438a@icu.ac.jp

1. Introduction

Drenjongke (also known as “Bhutia”, “Hloke” or “Sikkimese”) is a Tibeto-Burman language which is spoken by about 80,000 speakers in Sikkim, India and whose phonetic properties are understudied (see the green part in Figure 1.). Although Drenjongke is one of the official languages in Sikkim, the lingua franca languages in Sikkim are Nepali and English. Drenjongke is considered as endangered due to the decrease in the number of younger speakers. The literacy of Drenjongke is also not high because the Tibetan orthography is used for writing Drenjongke, which does not always succeed in representing the vernacular language.

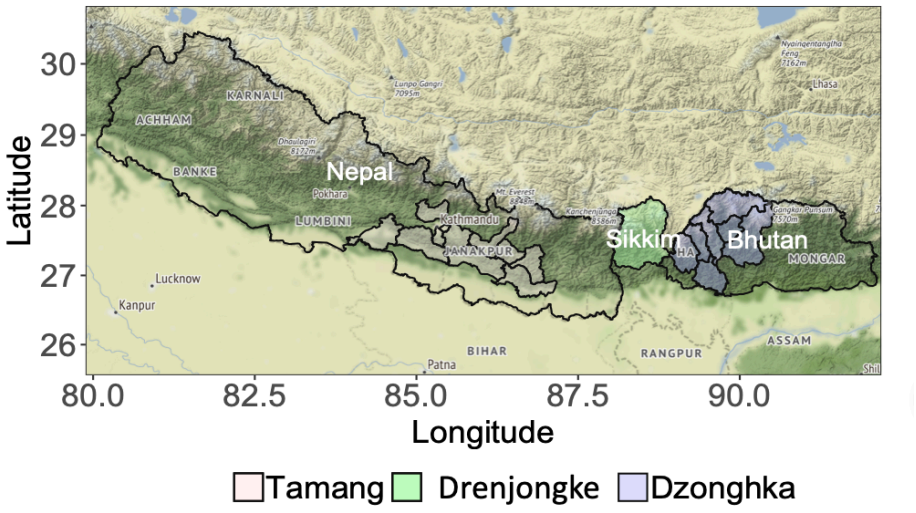


Figure 1. Map of the languages of the Himalaya

Descriptions of the language (vanDriem 2001, 2016; Yiliemni 2019) have reported a length contrast in some of the vowels of the phonological inventory of Drenjongke where ‘short’ vowels contrast with ‘long’ vowels. This contrast is involved in a variety of minimal pairs as exemplified in (1)

(1) Minimal pairs for the vowel contrast

a.	si	‘trouble, envy	si:	‘feel cool’	(Yiliemni 2019: 49)
b.	ka	‘order’	ka:	‘split’	(Yiliemni 2019: 49)
c.	ko	‘dig’	ko:	‘throw’	(Yiliemni 2019: 49)
d.	she	‘explain’	she:	‘know’	
e.	dru	‘boat’	dru:	‘six’	

However, what these studies also point out is that there is more to this contrast than a difference in vocalic duration. Both van Driem (2001, 2016) and Yiliemni (2019) report that only some of the vowels in the Drenjongke phonological inventory have this length contrast, and suggest that there is a complexity in the realization of this vowel contrast in relation to other acoustic differences such as vowel quality and the presence or absence of a glottal stop.

Although several impressionistic descriptions of this pattern are available, there is a lack of experimental studies examining acoustic properties of this vowel contrast. The current study offers a preliminary acoustic description of the production of the vowel length contrast by Drenjongke speakers in order to examine its acoustic realizations. After examining the durational cues, we looked at the different patterns of phonetic implementation exhibited by the ‘long’ vowels. Our findings, which are consistent with previous research results, confirm the complex nature of the contrast, and allow to identify a variety of patterns of phonetic realization for the ‘long’ vowel.

2. Methods

This study uses production data collected in March 2019 in Sikkim, India. The participants, eight native speakers of Drenjongke (5 male, 3 female), read a randomized list of words in a frame sentence with 5 repetitions. The list was made in order to include minimal pairs with a short vowel and its long counterpart for each vowel (e.g. [so] ‘tooth’ versus [so:] ‘save’).

The duration of each target segment was annotated using Praat (Boersma and Weenink 2018), and the extraction of measurements was automated with its scripting function. Statistical analyses were conducted using R (R core team 2017).

3. Results

We first investigated the annotated raw durations of the vowel segments. Our results are consistent with previous observations (van Driem 2001, 2016; Yiliemni 2019) that the vowel length contrast does not seem to be only based on a difference in vocalic duration.

The box plot in Figure 2 presents the distribution of the duration of short (left box) and long (right

box) vowels. In our aggregated data, the mean duration of all short vowels is 100 milliseconds (ms), while the mean duration for the long category is 110 ms, that is a durational ratio for the long/short vowel contrast of 1.1. Although the t-test indicated a significant difference between the two categories for the observed mean duration, the perceptual reality for native listeners of this difference (10 ms) is questionable. Moreover, what the box plots in Figure 2. also suggest is that considerable overlap exist in the distribution of the duration of the two categories.

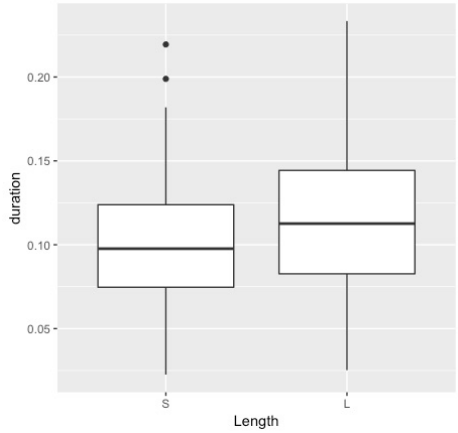


Figure 2. Distribution of the duration for short and long vowel categories

When looking at the same data organized by speaker (in Figure 3), we observe that the vowel length contrast is subject to inter-speaker variation. Each panel in Figure 3 illustrates the distribution of the vowel distinction based on orthography for each speaker. The plots in figure 3 show that the vowel length contrast have three ways of being implemented. Some speakers show no clear difference in duration between the short and the long categories (e.g. SIP052), and some others have a longer 'short' vowel (e.g. SIP049) or a longer 'long' vowel (e.g. SIP050). The duration results themselves do not offer a possibility of distinguishing short vowels from long vowels.

The distribution of the duration for the two categories is also of interest when we look at each pair separately. These pairs were examined because impressionistic studies have reported the presence of short versus long contrast in them. In Figure 4, we observe three different patterns: (a) word-pairs with no length contrast (e.g. A3-A4), (b) word pairs with a contrast with longer 'long' vowel (e.g. AMP15-16), or (c) word pairs with a contrast with a shorter 'long' vowel (e.g. MP21-22). The presence of inter-speaker variations in the realization of the length contrast, as well as the different patterns observed for the different pairs suggest that the vowel duration might not be the only acoustic correlate active for the vowel length contrast production.

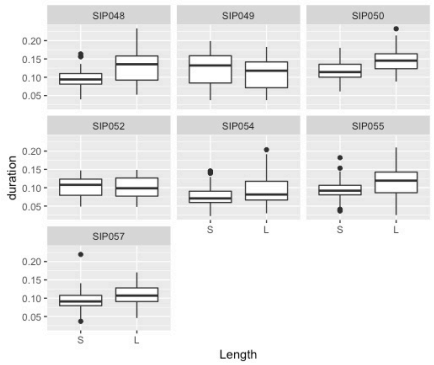


Figure 3. Distribution of vowel duration by speaker

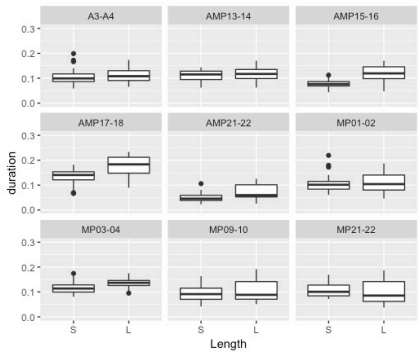


Figure 4: Distribution of vowel duration by pair

A further investigation of the recordings displays that there is no unique acoustic parameter that is responsible for the realization of the long vowel and that there is co-existence of several phonetic implementation patterns across the repetitions. Although short vowels consistently match the expected realization (i.e. a vowel with a short duration), we observe the following phonetic implementation patterns for long vowels:

(i) a longer duration of the vowel component when compared to its 'short' counterpart in the minimal pair (Figure 5.),

(ii) a short vowel followed by a consonant (Figure 6a.),

(iii) a difference in phonation: creaky voice (Figure 6b.),

(iv) a different vowel quality.

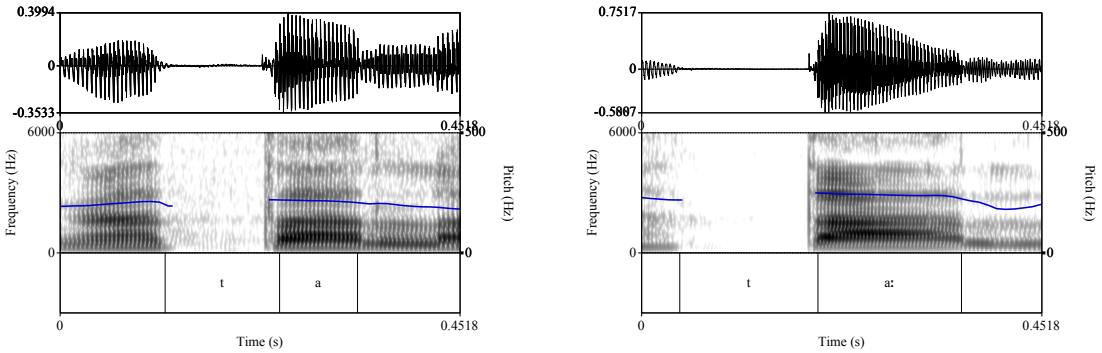


Figure 5. 'horse' [ta] vs. 'tiger' [ta:] minimal pair by speaker SIP071

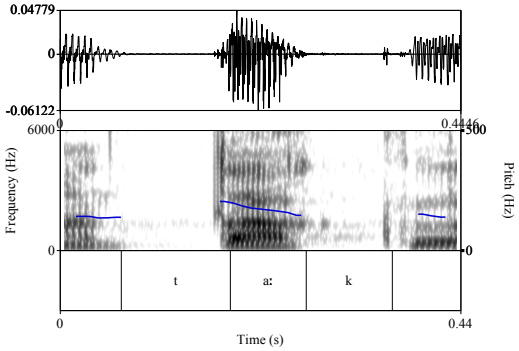


Figure 6a. ‘tiger’ /ta:/ pronounced as [tak] with a velar stop by SIP054

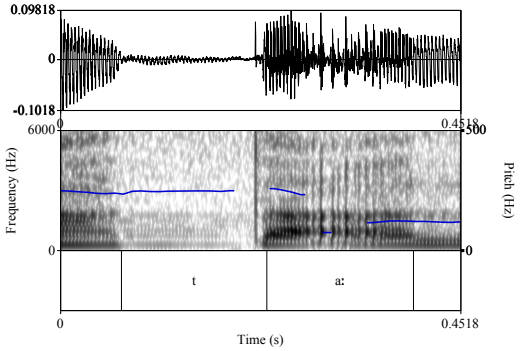


Figure 6b. ‘tiger’ /ta:/ pronounced as [tạ:] with creaky voice by SIP021

The spectrograms in Figure 5 and 6b illustrate three different types of phonetic implementation of the /ta/ ‘horse’ vs. /ta:/ ‘tiger’ [ta:] minimal pair. In Figure 5, we observe the expected realization of the short/long vowel contrast. On the left panel, the duration of the vowel for the short vowel is shorter than for its long counterpart in the right panel. In Figure 6a, the duration of the short vowel appears to be of similar duration with the short vowel in Figure 5, and the closure portion immediately following the vowel observed on the spectrogram suggests that the long vowel is realized as a short vowel followed by a stop. In Figure 6a, the consonant is a [k]. However, in other occurrences of long vowel we also observed glottal stops [ʔ], as well as [r] or [l] in this post-vocalic position. Figure 6b is an example of the long vowel realized with creaky voice, as shown by the glottal pulses in the spectrogram. Lastly, another type of phonetic implementation of the long vowel that is not illustrated in the spectrograms here is the difference in vowel quality. The pair A3-A4 /so/ ‘tooth’ vs. /so:/ ‘save’ was consistently realized as [so] for the short vowel, and [sɔ:] for its long counterpart. This is not surprising given that cross-linguistically vowel quality is a known correlate for vowel length contrast (Lehiste 1970, Maddieson 1984).

The four different patterns of phonetic implementation for the long vowel described above were not observed consistently. In fact, the patterns differ within an individual speaker (i.e. different realizations were observed through the five repetitions), between speakers (i.e. some speakers are more likely to lengthen or insert a consonant than others) and by item pairs (i.e. the same item pair may have various realizations). What our results suggest is that there is indeed a vowel length contrast in Drenjongke, but that the lengthening of the vowel duration is only one of the possible realizations and the contrast can be maintained using other cues.

4. Discussion and conclusion

This paper reports findings from the acoustic description of the vowel length contrast in Drenjongke. Although this language has been described in previous studies as having a contrast in terms of vowel

length opposing ‘short’ vowels to their ‘long’ counterparts, it appears that other acoustic correlates beyond duration need to be considered. Confirming previous studies (vanDriem 2001, 2016; Yiliemni 2019), several patterns of phonetic implementation was observed for the ‘long’ vowels only. There was no single phonetic parameter that is consistently observed across all the short vs. long vowel pairs; the acoustic realizations of the contrast between short and long vowels instead differ by speaker, between speakers, and by item pairs.

Research has shown that cross-linguistically, when a short-long contrast has a low durational ratio, other cues can be deployed to keep the distinction salient. This is for example the case in Norwegian where the duration of the vowel preceding duration of preceding vowel: Fintoft 1961). We suggest that this may also be the case in Drenjongke. When the vowel contrast is not saliently realized with a duration difference, the long vowel category utilizes other types of phonetic cues to maintain the contrast: a consonant can be inserted, the vowel is laryngealized, or the vowel can be differentiated. An interesting challenge is how to model this inter-speaker and inter-item variability.

Several questions arise from the findings of the present study. Firstly, do native speaker assimilate all the different realizations of the long consonant as the same phonemic category. Second, are native speakers able to make a perceptual distinction between the long and short categories. These questions will be addressed in a further study using perceptual experiments.

References

Boersma, P. and Weenink, D. (1992-2018) Praat: doing phonetics by computer. www.praat.org.

Fintoft, K. (1961) “The duration of some Norwegian speech sounds”, Phonetica 7, 19-39.

Lehiste, I. (1970) Suprasegmentals. Cambridge, MA.: MIT Press.

Maddieson, I. (1984) Patterns of Sounds. Cambridge University Press.

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

van Driem, G. (2001) Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language. Leiden: Brill.

van Driem, G. (2016) The phonology of Dränjoke. Manuscript.

Yiliemni, Juha (2019) A Descriptive grammar of Denjongke (Sikkimese Bhutia), Doctoral dissertation, University of Helsinki (with Sikkim University).

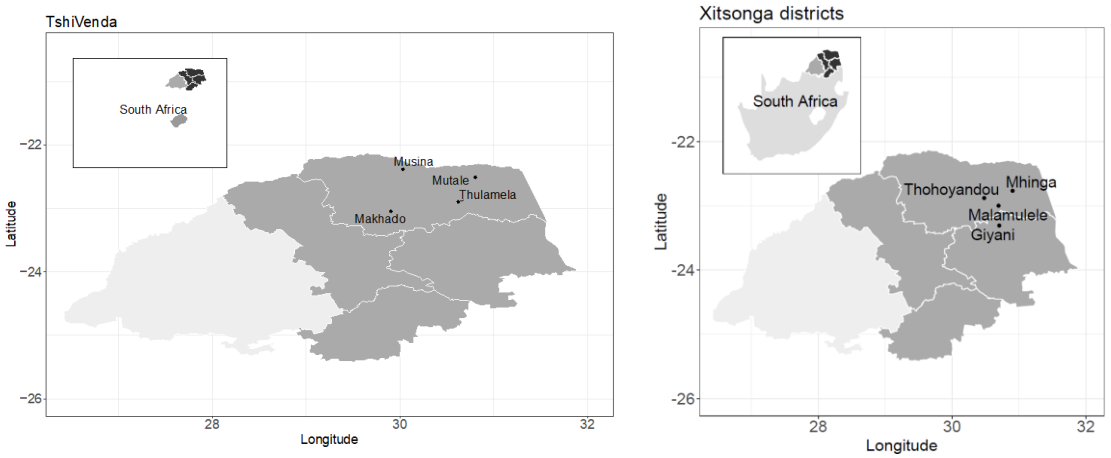
C2

Perception of a non-salient place contrast in Tshivenda by Xitsonga speakers

Michinori Suzuki(International Christian University)
Seunghun J. Lee(International Christian University, University of Venda)
michinori.suzuki.19@gmail.com, seunghun@icu.ac.jp

1. Introduction

Tshivenda is a southern Bantu language spoken in South Africa and Zimbabwe by about 1.3 million speakers.



a. Tshivenda speaking towns b. Xitsonga speaking towns

Figure 1: A map of Limpopo

Tshivenda is unique because it marks a place contrast between dentals and alveolars across various manners in the orthography using a carat sign underneath symbols for alveolars. Lee et al. (2018) reports acoustic characteristics of the place contrast and focuses on nasals due to qualitative differences found in the realization of plosives and laterals; for example, the orthographic 't' is produced as a palatal affricate and the orthographic 't̰' is produced as an alveolar. In Suzuki and Lee (2019), we report that the contrast between dentals and alveolars is retained in the production test, but the difference is not robustly perceived by native speakers of Tshivenda. This finding was interpreted as perceptual merger in the non-salient coronal place contrast, also known as near merger (Yu 2007).

In this study, we address the issue of perception of non-salient contrast by looking into perception data from Xitsonga speakers who don't have the nasal place contrast, but nonetheless share the linguistic sphere with Tshivenda speakers. Informal inquiries to Xitsonga speakers yielded responses that they can distinguish the place contrast, presumably because sharing the linguistic

sphere allowed them to observe the production of the sounds. Even so, an experimental study was necessary to verify these impressionistic claims to understand non-native speakers' perception of the non-salient contrast.

Xitsonga speakers are selected for this study due to various reasons. Xitsonga speakers share the geographical area with Tshivenda speakers. During the apartheid period, the homeland for Tsonga people and the homeland for Venda people were next to each other. In remote rural areas, intermarriage between these two groups is common. In modern day Thohoyandou, Limpopo, contacts between these two groups are frequent so that many speakers can understand the basic of each other's language. This geographical proximity and close social structure make a perception study of these two languages worthwhile because other external factors can be excluded in understanding the perception of non-salient contrast by speakers of other languages.

Perception results reveal that Xitsonga speakers do not perceive the difference between dental and alveolar place contrast (even though some informally claimed that they can do so). The findings suggest that geographical proximity as well as familiarity with a language do not result in a perception of a non-salient contrast. Furthermore, the findings imply that non-salient phonological contrast may only be perceivable only if the grammar encodes the contrast.

2. Experiment

2.1. Participants and Procedure

A perception task was conducted in Thohoyandou, Limpopo, South Africa in November 2018. Xitsonga speakers who were familiar with the orthographic representation of the dental and alveolar contrast were recruited. Eleven native speakers of Xitsonga were asked to identify whether a token in a frame sentence begins with a dental or an alveolar. Two forced choices were presented using standard Tshivenda orthography using Superlab version 5.0 (Abboud 2013). Participants made decisions using two keys on a keypad that was directly connected to a MacBook Air using a USB connection. After a short practice session, participants made judgements for 120 items selected from the production data by four speakers.

2.2. Stimuli

Stimuli for the perception test were 3 pairs of minimal pairs with the dental versus alveolar contrast in (1). Recordings of the 6 items from 6 previously recorded participants were selected for the perception test. At the time of the recording, a post-test confirmed that all the speakers who produce the stimuli were using these words in their everyday life.

(1) Stimuli

a. n̪ínga ‘punch’ nínga ‘hit sideways’

b. n̪ànga ‘flute’ nánga ‘choose’

c. n̪ènga ‘sneak away’ nènga ‘sneak’

2.3. Analysis

Results from the psychological software Superlab 5 were analyzed using a d-prime value. Perception tests yield four types of responses that need to be considered. For example, when a participant responds with ‘dental’, the response may come from a dental stimulus (called ‘hit) or from an alveolar stimulus (called ‘false alarm’). In that same scenario, when a participant responds with ‘alveolar’, the response may come from an alveolar stimulus (called ‘correct rejection’) or from a dental stimulus (called ‘miss’). The d-prime value is calculated by subtracting the z-score of ‘false alarm’ responses from the z-score of ‘hit’ items. When two sounds are categorically distinguished by participants, the d-prime value reaches around 3.

One participant only provided alveolar responses. We deemed that this participant did not understand the purpose of the experiment, and the responses were excluded from further data analysis.

3. Results

The results show that Xitsonga listeners do not distinguish dentals from alveolars based solely on acoustic recordings. Although the higher the d-prime is, the more discernability of two items, d-prime values by Xitsonga listeners were rather low as shown in table 1. The lowest d-prime value is -0.17, and the highest value is 0.45. The overall distribution of d-prime value suggests that Xitsonga speakers cannot distinguish the nasal place contrast between dentals and alveolars. The c-value shows a tendency in their responses. Higher c-values indicate that a participant tends to offer alveolar responses. Given the low c-value, Xitsonga speakers have tendencies to respond most items as dentals.

Table 1: D-prime and C values

Listeners	D-prime	C
1	**-0.179555829**	-0.163569188
2	0.125355438	0.147750675
3	-0.030370821	-0.216648952
4	0.178345967	0.036488363

5	-0.115820944	0.037020384
6	-0.011685931	0.131504312
7	0.05268462	-0.02634231
8	0.326541345	0.047157722
9	**0.458182462**	-0.103429884
10	-0.074980604	-0.347830164

4. Discussion

It is well known that perceiving contrasts that are not present in one's dominant or native language is not an easy task. However, it is not always obvious whether speakers of languages that share the same linguistic sphere will share the same difficulty. Xitsonga speakers and Tshivenda speakers not only live in a nearby area, but also share various aspects of their life suggesting a closer than usual relationship both linguistically or non-linguistically. Such a close relationship would predict that non-native contrast in Tshivenda may not be difficult to perceive by speakers of Xitsonga.

The findings in this study, however, show that non-salient contrast in Tshivenda that is not present in Xitsonga is not perceivable. D-prime results by Xitsonga speakers show that many listeners were guessing rather than discriminating any differences between dental and alveolars. Moreover, when they have to make a judgment, Xitsonga speakers displayed inclination to judge a stimulus as a dental sound rather than an alveolar sound; possibly due to the marked status of the dental nasal sound. In other words, when Xitsonga speakers face a difficult task of mapping an audio stimulus to a category of sound, they opted for the marked dental rather than the unmarked alveolar.

In Suzuki and Lee (2019), Tshivenda speakers display rather low d-prime values for an alleged native contrast (ranging from -0.34 to 1.62), which suggests that near merger (Yu 2007) is an ongoing process in the phonology of Tshivenda. This finding shows that Tshivenda speakers already have difficulty perceiving the nasal place contrast even though they could produce the difference by advancing the tongue and producing an interdental nasal. The results in this paper show that a contrast that is experiencing near merger is also nearly impossible to be perceived by speakers of other languages when the contrast is not present in those languages.

References

Abboud, H. (2013). Superlab version 5.0. San Pedro, California. Cedrus.

Fitzpatrick, J. and Wheeldon, L. (2000) Phonology and phonetics in psycholinguistic models of speech perception. In: Burton-Roberts, N. et al. (eds.) Phonological Knowledge: Conceptual and Empirical Issues. Oxford University Press. (pp. 131-160).

Guthrie, M. (1967–1971) Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages. Farnborough: Gregg Press.

Lee, S.J., Tshithuke, S., Suzuki, M. (2018) An acoustic study of dental vs. alveolar contrast in Tshivenḓa nasals. The 156th Linguistics Society of Japan.

Lindblom, B. (1963) On vowel reduction. Technical Report 29 The Royal Institute of Technology, Speech Transmission Laboratory. Stockholm, Sweden.

Poulos, G. (1989) A linguistic analysis of Venda. Grammar, Via Africa.

Radzhadzhi, M. A. (2002) Nasal assimilation and related processes in Tshivenda: a linear and non-linear phonological analysis. MA thesis, University of Stellenbosch

van Warmelo, N. J. (1995) Venda Dictionary. Hippocrene Books.

Suzuki, M. Lee, S.J. (2019) Production and perception of dental vs. alveolar contrast in Tshivenda. Proceedings of the International Congress of Phonetic Sciences 2019.

Yu, A. (2007). Understanding near mergers: The case of morphological tone in Cantonese. Phonology, 24(1), 187-214. doi:10.1017/S0952675707001157

C3

日本語母語話者によるロシア語の無声舌頂阻害音の知覚

VAKHROMEEV ANATOLII（上智大学）

1. はじめに

本研究はロシア語無声舌頂阻害音の聴覚的調査を通じて，非母語の音声知覚に言語学的視点からアプローチしたものである．本研究の重要な課題の 1 つはロシア語を学習したことがない日本語母語話者がどのような知覚の状態からロシア語の学習を開始するかということを調べることである．そのために，本研究ではロシア語を一切学習していない日本語母語話者とロシア語母語話者の知覚を比較する．

本研究が扱うロシア語の無声舌頂阻害音音類は /t, tʲ, t͡s, t͡ɕ, s, sʲ, ʂ, ɕ/ である．日本語母語話者がロシア語を習得する際，閉鎖音音素においては，新たに /tʲ/ [tʲ]（破擦性を伴うため [t͡ɕ] という精密表記ができる）を習得し，/t͡ɕ/ と /tʲ/ の弁別を取得するという学習課題がある．さらに，摩擦音音素においては，/sʲ/ [sʲ] と /ʂ/ [ʂ] を新たに習得し，/sʲ, ʂ, ɕ/ の弁別を習得する必要がある．日本語母語話者にとって /t͡ɕ/ と /tʲ/ の発音の弁別が困難であるといわれている（城田 1979，神山 2012）．ヴァフロメーエフ 2018 では産出においてこの 2 音素の混同が観察された．知覚においてもこの 2 音素の混同が予測される．日本語母語話者によるロシア語の摩擦音音素の知覚に関してはこれまで未調査であるが，本研究では対立に関わる音響素性の数が同じであるため /ʂ, sʲ, ɕ/ が同程度に知覚において混同されるという仮説を立てた．

2. 方法

2.1. 被験者

調査に参加した日本語母語話者は全員ロシア語を学習した経験がない，ロシア語の未学習者である．L2 知覚への影響をなるべく均質なものにするために，協力者の出身地および言語学習歴を可能な限り統制した．日本語母語話者は 1 名を除いて関東出身であり，残りの 1 名は北海道出身であった．日本語母語話者全員の第 1 外国語は英語であった．日本語母語話者は 2 名を除いてすべての被験者は朝鮮語専攻の大学 1 年生および 2 年生で，第 2 外国語が韓国語であった．残りの 2 名は日本語専攻で韓国語が第 2 外国語である大学院生と，ドイツ語専攻でドイツ語が第 2 外国語である大学 4 年生であった．日本語母語話者は全員 20 代前半であった．調査に参加した 7 名のロシア語母語話者は全員 20 代で，大学生および大学院生であった．

2.2. 調査に用いた刺激音

調査には 2 種類の無意味語を用いた（一部の生成した単語はロシア語において有意味語として存在する）．異なる単語の対＝異語対と同じ単語の対＝同語対である．同語対の場合，下に述べた子音の中から，同じ子音が使われ，異語対の場合，異なる子音が使われた．例えば，/tak/—/t͡sak/ は異語対であり，/tak/—/tak/ は同語対である．刺激音の無意味語は，調査対象の子音 /t/，/tʲ/，/t͡s/，/t͡ɕ/，/s/，/sʲ/，/ʂ/，/ɕ/ と母音 /a, o/ および /k/ から編成した．語頭の部門の場合，調査対象の子音および子音 /m, n/，/r, l/，/b, v/ が語頭にあり，母音 /a, o/ が後続し，/k/ が語末にある単語の対を用いた．語末の部門の場合は /k/ が語頭にあり，

母音 /a,o/ が後続し，語末に調査対象の子音がある対を用いた．同語対と異語対は表 1 の同一セル内部にある無意味語の組み合わせを網羅的に用いて編成した．

表 1: 調査対象の無意味語対

語頭の部門		語末の部門	
_/ak/	_/ok/	/ka/_	/ko/_
/tak/	/tok/	/kat/	/kot/
/tʲak/	/tʲok/	/katʲ/	/kotʲ/
/t͡sak/	/t͡sok/	/kat͡s/	/kot͡s/
/t͡ɕak/	/t͡ɕok/	/kat͡ɕ/	/kot͡ɕ/
/sak/	/sok/	/kas/	/kos/
/sʲak/	/sʲok/	/kasʲ/	/kosʲ/
/ʂak/	/ʂok/	/kaʂ/	/koʂ/
/ɕak/	/ɕok/	/kaɕ/	/koɕ/

刺激音は Johnson (2003: 61-63)に述べられた方法を採用し，聴覚的区別を困難にするために，刺激音にはブロードバンドノイズを 0 dB SNR でミックスした．そして，予めノイズのミックスした刺激音の WAV ファイルを刺激提示ソフトウエア SuperLab 5 の中で再生した．

2.3. 調査の手順

実験的調査は 2 つの部門から編成した．1 つめの部門では分析対象の子音が単語の語頭にある単語対を被験者に聞かせ，2 つめの部門では単語の語末にある単語対を聞かせた．前者の部門を一貫して，「語頭の部門」と呼ぶことにし，後者を「語末の部門」と呼ぶことにする．この「語頭の部門」と「語末の部門」では，それぞれの語頭の位置と語末の位置における音素の知覚的データを収集した．実験的調査の部門の順番は常に，語頭の部門が最初で，語末の部門が最後であった．さらに，この 2 つの部門の前に，インストラクションと練習を行った．語頭と語末の部門を合わせて，1 名の被験者に 204 個の異語対および 84 個の同語対を聞かせた．

インストラクションにおいては，調査の内容や手順および機材の使い方について被験者に画面上の説明を読んでもらった．ロシア語母語話者にロシア語のインストラクションを示し，日本語母語話者に日本語でのインストラクションを示した．インストラクションの重要な点として，語頭の部門の場合は語頭の子音が、語末の部門の場合は語末の子音が注目するポイントである．さらに，母音には /a/ と /o/ があると伝えた．この説明はインストラクションの部門および，それぞれの語頭と語末の部門を始める前に読んでもらった．

刺激音を聞かせる際に，ダイナミック・ステレオ・ヘッドホン SONY MDR-CD900ST を用いた．音量はすべての被験者に対して一定の音量に指定した．なお，被験者の判断データは SuperLab 5 の専用のレスポンス・ボックスを用いて収集した．

2.4. 知覚的データ分析方法および解釈方法

被験者の判断のデータを，判断の正誤（異語対に対する「違う」という判断および，同語対に対する「同じ」という判断は正答であり，異語対に対する「同じ」という判断および，同語対に対する「違う」という判断は誤答である）の割合を，それぞれの当該子音の対別にプロットし，観察した．正答率の低い対は知覚的距離が近く、類似性が高いと解釈し、正答率が高い対は知覚的距離が遠く、類似性が低いと解釈した。さらに，異語対間で分散分析と多重比較を行い，子音の対間の判断の正誤の有意差（有意水準 1%）を確認した．なお，対の順番による顕著な差異は認められなかったために，順番が異なる対を観察する場合に，片方の順番に統一した．つまり，例えば，/sak/—/ʂak/ という対と /ʂak/—/sak/ という両方の対に関する判断は「sak—ʂak」という風にまとめて分析対象とした．さらに，後続母音による差異も認められなかったために，後続母音が /a/ と /o/ の対を分けていない．

3. 結果と考察

この節では，被験者の無声舌頂阻害音の正答率の分析結果について述べる．

3.1. 閉鎖音音素

図 1 に示すロシア語母語話者の場合，語頭では，正答率が低いのは /t͡ɕ/ と /tʲ/ および /ts/ と /tʲ/ である．さらに，/t͡ɕ/ と /t͡s/ および破擦的開放を伴う /t͡ɕ, t͡s, tʲ/ と /t/ は知覚的な類似性が低いことが分かった．語頭と語末で共通するのは，/t/ と /tʲ/，/t͡ɕ/ と /t͡s/，/t/ と /t͡ɕ/ の知覚的距離が比較的遠く，/t͡ɕ/ と /tʲ/ および /t͡s/ と /tʲ/ の知覚的距離が近いことである．一方，語頭と大きく異なるのは /t/ と /t͡s/ の知覚的距離が近いことである．ただし，語末の場合，子音の対間に有意な差異は認められなかった．

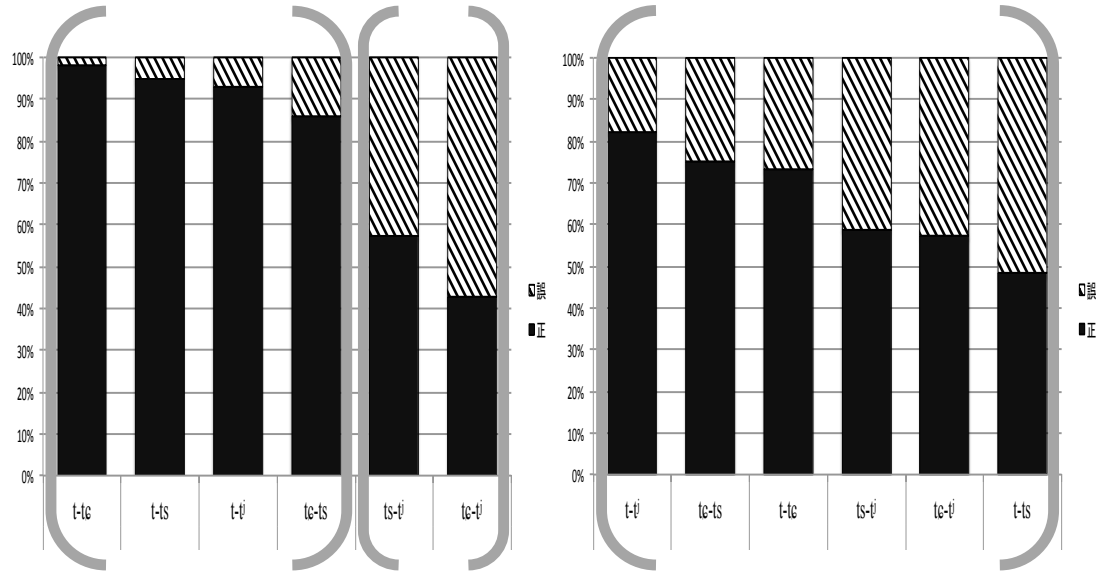


図 1: ロシア語母語話の閉鎖音対の区別の正答率: 語頭（左）および語末（右）．それぞれの縦棒は正答率の平均値を表す．棒グラフをくくる括弧は分散分析で有意差の認められたクラスを示す．

図 2 に示す日本語母語話者の場合，語頭では，/tɕ/ と /tʲ/ が他の対より，正答率が極めて低いことから，混同が生じるほど（区別されないほど）類似性が高いことが分かった．また，/tɕ/ と /tʲ/ と有意な差はあるが /ts/ と /tʲ/ および /tɕ/ と /ts/ の正答率も比較的低い．破擦的開放を伴う 3 音素の /tɕ, ts, tʲ/ と /t/ は正答率が高く，類似性が低いことが分かった．語末の場合，語頭と異なるパターンが観察された．語末で全体的に低い正答率が見られ，子音の区別が大きな困難を伴うことが示唆された．語末の /t/ と /tɕ/ は他の対と有意な差があるが，それ以外の子音音素の対には有意な差が認められない．

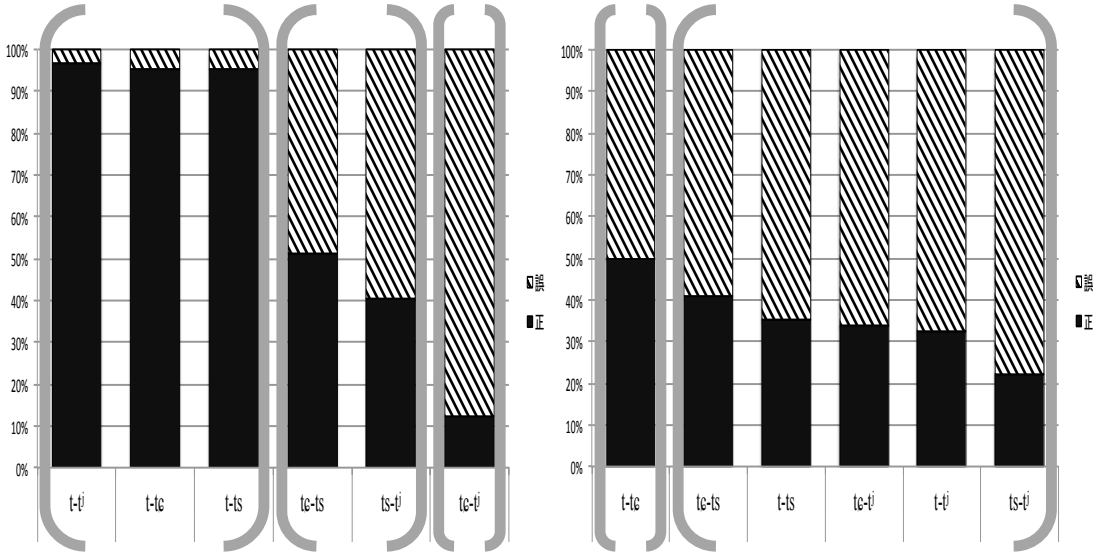


図 2: 日本語母語話の閉鎖音対の区別の正答率: 語頭（左）および語末（右）. それぞれの縦棒は正答率の平均値を表す. 棒グラフをくくる括弧は分散分析で有意差の認められたクラスを示す.

3.2. 摩擦音音素

ロシア語母語話者の場合，図 3 に示す通り，語頭では，/s/ と /sʲ/ の正答率が他の対より有意に低いことが分かった．それ以外の対の間では有意差は認められなかった．また，語末では，子音の対の間で有意差は認められなかった．

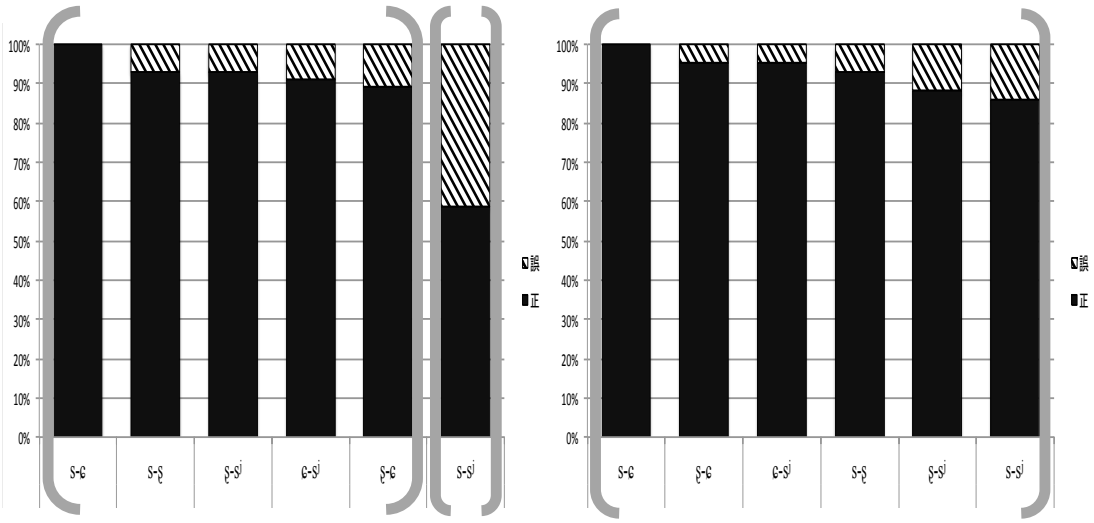


図 3: ロシア語母語話の摩擦音対の区別の正答率: 語頭(左)および語末(右). それぞれの縦棒は正答率の平均値を表す. 棒グラフをくくる括弧は分散分析で有意差の認められたクラスを示す.

図 4 に示す日本語母語話者の場合，語頭では，/ʂ/ と /ɕ/ の正答率がもっとも低い．この対の正答率から，日本語母語話者の場合，/ʂ/ と /ɕ/ はほとんど区別されないことが明らかになった．/ʂ/ と /ɕ/ の次に知覚的距離が近いのは /s/ と /sʲ/ である．もっとも正答率が高いのは /s/ と /ɕ/ および /s/ と /ʂ/ である．語末では，語頭と同じように，/ʂ/ と /ɕ/ の正答率がもっとも低いために，類似性が高いことが分かった．上に述べたことから，日本語母語話者の知覚では，/ɕ/ と /ʂ/ の混同が生じることが明らかになった.

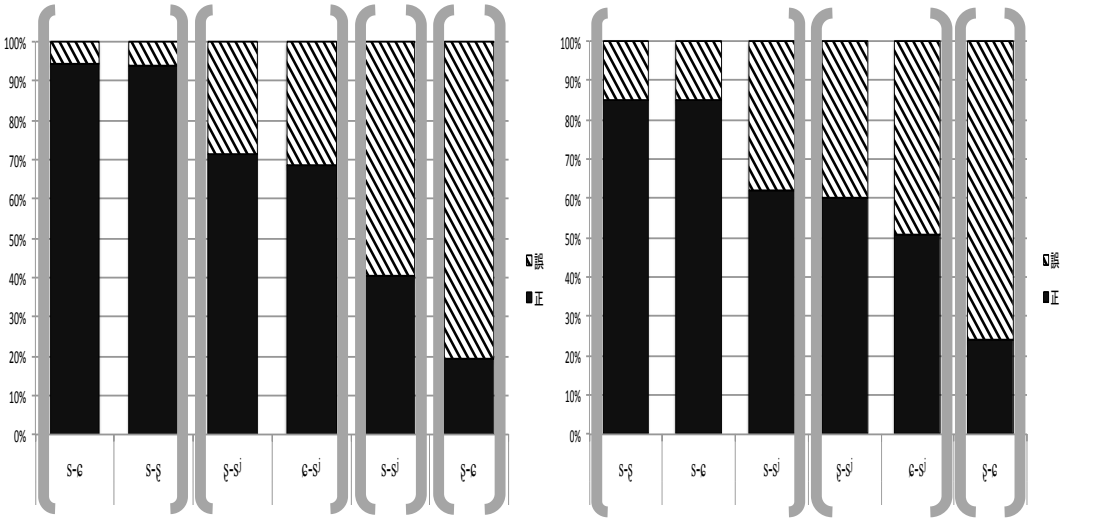


図 4: 日本語母語話の摩擦音対の区別の正答率: 語頭(左)および語末(右). それぞれの縦棒は正答率の平均値を表す. 棒グラフをくくる括弧は分散分析で有意差の認められたクラスを示す.

4. 結論

ロシア語を学習していない, 日本語を母語とする被験者の場合, 閉鎖音においては, /t͡ɕ/ と /tʲ/ が予測通りに混同されることが分かった．一方，摩擦音においては，/ʂ, sʲ, ɕ/ が同程度に知覚において混同されるという仮説は棄却された．/ɕ/ と /ʂ/ は区別されないほど知覚的に類似性が高いことから，この 2 音素が同じ範疇であると結論付けた．

従って，L2 ロシア語の学習の出発の段階で，日本語母語話者であるロシア語学習者の知覚では, 閉鎖音において, /t/, /t͡s/, /t͡ɕ, tʲ/ の 3 つの範疇の存在が示唆され, 摩擦音において, /s/，/sʲ/，/ɕ, ʂ/ の 3 つの範疇の存在が示唆された．この範疇化が学習と共にどう変化していくかについて調べることは今後の課題である．

参考文献

Johnson, Keith. 2003. *Acoustic and Auditory Phonetics*, Second edition. Blackwell Publishing.

城田俊 (1979)『ロシア語の音声 ： 音声学と音韻論』, 風間書房.

神山孝夫 (2012)『ロシア語音声概説』, 研究者.

ヴァフロメーエフ A. (2018) 日本語母語話者による L2 ロシア語の産出における /t͡ɕ/ と /tʲ/ の混同,『ロシア語研究』第 28 号.

C4

ベトナム語北部方言における「短母音＋舌背音」の韻について

山岡翔（京都大学大学院文学研究科・日本学術振興会特別研究員）
sho.yamaoka@gmail.com

1. はじめに

ベトナム語北部方言では、「短母音＋舌背音」からなるような一連の韻が存在し、現在以下のように記述・分析されることが一般的である。

表 1: ベトナム語北部方言における「短母音＋舌背音」からなる韻の記述と分析①[1]

/iŋ/, /ik/ → [i̯ŋ:], [i̯c:]	/ɯŋ/, /ɯk/ → [ɯ̆ŋ:], [ɯ̆k:]	/uŋ/, /uk/ → [ŭʷŋ͡m:], [ŭʷk͡p:]
/eŋ/, /ek/ → [e̯ŋ:], [e̯c:]	/ɤŋ/, /ɤk/ → [ɤ̆ŋ:], [ɤ̆k:]	/oŋ/, /ok/ → [ɤ̆ʷŋ͡m:], [ɤ̆ʷk͡p:]
/ɛŋ/, /ɛk/ → [ɛ̯ŋ:], [ɛ̯c:]	/ăŋ/, /ăk/ → [ʌ̆ŋ:], [ʌ̆k:]	/ɔŋ/, /ɔk/ → [ʌ̆ʷŋ͡m:], [ʌ̆ʷk͡p:]

これはつまり、一連の韻の音韻的区別は末子音ではなく母音にあるとする分析である。一方で、一連の韻の音韻的区別を母音ではなく末子音が担っているとする以下の表 2 のような分析も可能である。しかし、どちらの分析が妥当なのかはいまだ明らかにされておらず、現在は経済性や分布の均一性の観点から専ら表 1 の分析が好まれるのが現状である。

表 2: ベトナム語北部方言における「短母音＋舌背音」からなる韻の記述と分析②[2]

/ɯɲ/, /ɯc/ → [i̯ŋ:], [i̯c:]	/ɯŋ/, /ɯk/ → [ɯ̆ŋ:], [ɯ̆k:]	/ɯŋ͡m/, /ɯk͡p/ → [ŭʷŋ͡m:], [ŭʷk͡p:]
/ɤɲ/, /ɤc/ → [e̯ŋ:], [e̯c:]	/ɤŋ/, /ɤk/ → [ɤ̆ŋ:], [ɤ̆k:]	/ɤŋ͡m/, /ɤk͡p/ → [ɤ̆ʷŋ͡m:], [ɤ̆ʷk͡p:]
/ăɲ/, /ăc/ → [ɛ̯ŋ:], [ɛ̯c:]	/ăŋ/, /ăk/ → [ʌ̆ŋ:], [ʌ̆k:]	/ăŋ͡m/, /ăk͡p/ → [ʌ̆ʷŋ͡m:], [ʌ̆ʷk͡p:]

そこで本稿では、音響的観点から一連の「短母音＋舌背音」の韻は以下の表 3 のように音声表記・音韻分析されるべきであることを主張する。

表 3: 本稿の主張する、ベトナム語北部方言における「短母音＋舌背音」からなる韻の記述と分析

/iɲ/, /ic/ → [i̯ŋ:], [i̯c:]	/iŋ/, /ik/ → [i̯ŋ:], [i̯k:]	/iŋ͡m/, /ik͡p/ → [i̯ʷŋ͡m:], [i̯ʷk͡p:]
/ɜɲ/, /ɜc/ → [ɛ̯ŋ:], [ɛ̯c:]	/ɜŋ/, /ɜk/ → [ɛ̯ŋ:], [ɛ̯k:]	/ɜŋ͡m/, /ɜk͡p/ → [ɛ̯ʷŋ͡m:], [ɛ̯ʷk͡p:]
/ăɲ/, /ăc/ → [ă̯ŋ:], [ă̯c:]	/ăŋ/, /ăk/ → [ă̯ŋ:], [ă̯k:]	/ăŋ͡m/, /ăk͡p/ → [ă̯ʷŋ͡m:], [ă̯ʷk͡p:]

2. 問題の所在

本節では、「短母音＋舌背音」からなる韻にまつわる問題点を改めて整理する。

---

[1] Cao Xuân Hạo (2007: 95), Đoàn Thiện Thuật (2007: 234) の表記に、発表者が分かりやすさのため短母音化・長音化の記号を足したもの。ただし、[i̯], [ʷ] は子音の二次的調音ではなく、母音・末子音間に現れるごく短いわたりを表す。

[2] Đoàn Thiện Thuật (2007: 255) を参考に発表者が作成した。

2.1. 「短母音＋舌背音」からなる韻の種類

表 4:「短母音＋舌背音」からなる韻の分類と、従来の音声的記述

開口度	短母音＋硬口蓋音の韻	短母音＋軟口蓋音の韻	短母音＋両唇軟口蓋音の韻
小	[ɨ̆ɲ:], [ɨ̆c:]	[ŭŋ:], [ŭk:]	[ŭʷŋ͡m:], [ŭʷk͡p:]
中	[ə̆ʲɲ:], [ə̆ʲc:]	[ɤ̆ŋ:], [ɤ̆k:]	[ɤ̆ʷŋ͡m:], [ɤ̆ʷk͡p:]
大	[ă̆ʲɲ:], [ă̆ʲc:]	[ă̆ŋ:], [ă̆k:]	[ă̆ʷŋ͡m:], [ă̆ʷk͡p:]

ベトナム語北部方言における「短母音＋舌背音」からなる韻は上に示したように、末子音について硬口蓋音・軟口蓋音・両唇軟口蓋音をもつものの 3 系列、さらに母音の開口度について小中大の 3 系列の、計 9 種類が存在する。硬口蓋音系列の韻の母音部分は非前舌から前舌へ、両唇軟口蓋音系列の韻の母音部分は非円唇から円唇かつ非後舌から後舌へと素早く変化するようなわたり [ʲ], [ʷ] をもつ。

2.2. 「短母音＋舌背音」からなる韻の二種類の分析方法

これら一連の「短母音＋舌背音」からなる韻の分析方法は長きにわたって議論がなされてきた[3]。各先行研究が行っている一連の韻の分析方法は多岐にわたるが、本稿では便宜上これらの方法を「各韻の音韻的区別を母音に認めるか末子音に認めるか」によって二分し、どちらの分析が妥当かを比較するような方向で論を進める。

まずひとつめの分析は、当該の各韻の母音を別音素と捉え、末子音はすべて同一音素である軟口蓋音の異音と考える表 1 のような分析である（Cao Xuân Hạo 2007: 95 など）。なお、この分析において基底から表層へと派生させるためには、以下のような一連の規則が必要となる（Cao Xuân Hạo 2007: 95）。このような分析を以下、母音説と呼ぶ。

(1)　「短母音＋舌背音」からなる韻の母音説における派生規則

	iŋ	eŋ	ɛŋ	uŋ	oŋ	ɔŋ
短母音化	ĭŋ:	ĕŋ:	ɛ̆ŋ:	ŭŋ:	ŏŋ:	ɔ̆ŋ:
口蓋化・唇音化	ĭɲ:	ĕɲ:	ɛ̆ɲ:	ŭŋ͡m:	ŏŋ͡m:	ɔ̆ŋ͡m:
母音の異化	ɨ̆ɲ:	ə̆ɲ:	ă̆ɲ:	ŭŋ͡m:	ɤ̆ŋ͡m:	ă̆ŋ͡m:
わたりの付加	ɨ̆ʲɲ:	ə̆ʲɲ:	ă̆ʲɲ:	ŭʷŋ͡m:	ɤ̆ʷŋ͡m:	ă̆ʷŋ͡m:

ふたつめの分析は、各韻の母音は同一音素であり、代わりに末子音に硬口蓋音・軟口蓋音・両唇軟口蓋音の 3 つの系列の区別があるとする表 2 のような分析である（Nguyễn Phan Cảnh: 1964 など）。この分析を以下、末子音説と呼ぶ。

これらふたつの分析方法のうち現在より一般的なのは母音説であるが、その理由としては末子音の音素の数が少なくて済むこと、そして韻の分布（母音＋末子音の組み合わせの種類）が音素によらずほぼ均一になることなどの利点のためである。

[3] 各先行研究の分析方法の分類や議論の内容については Cao Xuân Hạo (2007: 88–91), Đoàn Thiện Thuật (2007: 237–266) などが詳しい。

そこで本稿では、「短母音＋舌背音」からなる韻を音響的に分析し、母音説と末子音説のどちらの分析が妥当なのかを改めて検討する。

3. データの収集および分析

本節では、「短母音＋舌背音」からなる韻の音響分析に用いるデータの収集および分析について述べる。

3.1. 収集方法

ハノイ方言話者 1 名にターゲットとなる韻を含む語彙のリストを読み上げてもらった。ひとつの韻につき、3 回ずつ読み上げている。収録は比較的静かな部屋[4]で行い、Tascam DR-40 の内臓マイク（コンデンサー、全指向性）を用いて 44.1kHz で標本化、16bit で量子化した WAV ファイルで出力した。

3.2. 分析方法

praat のバージョン 6.0.08（Boersma & Weenink 2015）にて各語の母音部分のフォルマントの持続部分を選択し、そのなかから長母音は 10 点（選択箇所の開始地点を 0%、終了地点を 100% としたとき、5%, 15%, …, 95% となる 10 点）、短母音は 5 点（10%, 30%, …, 90% となる 5 点）のフォルマントを取得し、平均をとった。ただし、外れ値は除外している。

4. 分析結果

本節では、前節で説明したフォルマント分析の結果について見る。

4.1. ハノイ方言の母音体系

分析結果の前に、まずハノイ方言の母音体系について概観する。

表 5 ハノイ方言の母音体系

		前舌	中舌		後舌
			長母音	短母音	
		非円唇			円唇
単母音	高母音	/i/	/ɨ/		/u/
	中段母音	/e/	/ə/	/ɜ̆/	/o/
	低母音	/ɛ/	/a/	/ă/	/ɔ/
二重母音	下り二重母音	/iə/	/ɨə/		/uə/

中舌母音にのみ長短の対立があるが、この長短の対立は母音の持続時間だけでなく、同時に末子音の持続時間にも影響する。つまり、「母音の長短」というより、「母音と末子音の長さの比率の大小」といった方が正確である。なお、短母音は閉音節にしか現れない。

[4] ただし、部屋は全面石造りでやや反響しやすい構造になっている。また、室内の家電製品や屋外からのノイズもみられるが、コンサルタントの音声のレベルに比べて非常に小さいので、分析上大きな問題はないと思われる。

(2)　母音の長短と末子音の関係

長母音＋末子音　/VC/　→ [V:Č]

短母音＋末子音　/V̆C/　→ [V̆C:]

また、中舌高母音 /ɨ/ は長短の対立するペアをもたないが、この母音は音声的環境により長短が変化する。

(3)　中舌高母音 /ɨ/ の変異

/ɨ/　→ [ɨ̆] / _ [+ consonantal]

→ [ɨ:] / elsewhere

なお、中舌中段母音の /ə/, /ɜ/ は母音の長短についてだけでなく、音色についてもやや異なっている。清水（2007: 28–30）は /ə/, /ɜ/ のフォルマント値に重なりがないことのほか、閉音節における母音 /ə/ を短くしても、また母音 /ɜ/ を長くしても、母語話者はもとの母音のままであると知覚する傾向があることを述べている。

4.2.　ハノイ方言の「短母音＋舌背音」からなる韻の分析

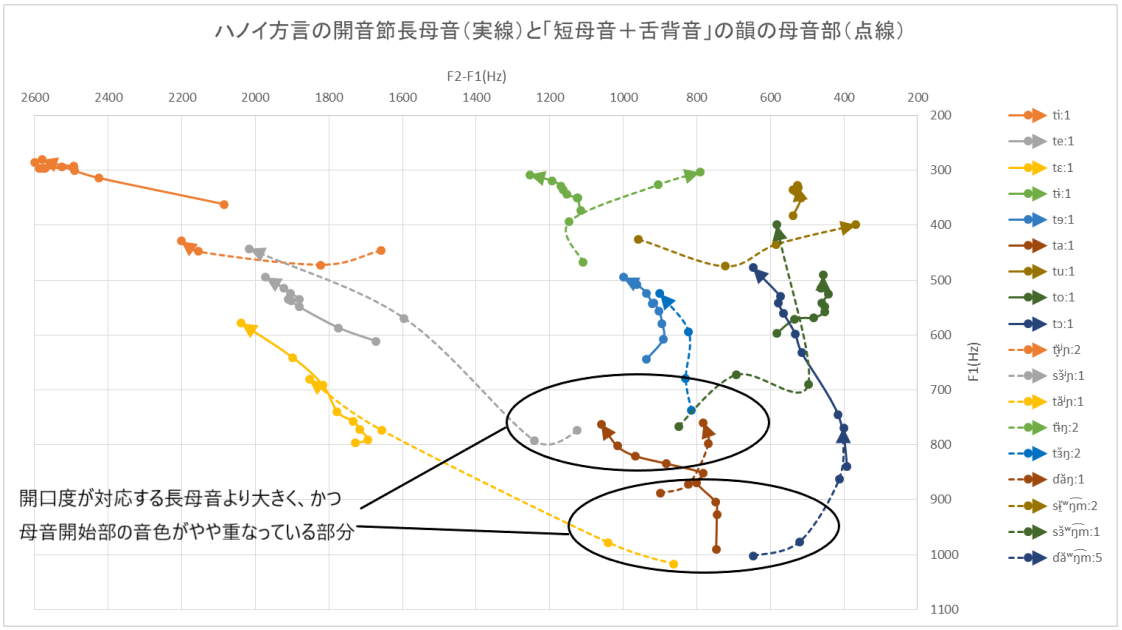


図 1: ハノイ方言の「短母音＋舌背音」からなる韻と開音節単母音の韻における母音のプロット

上図はハノイ方言の「短母音＋舌背音」からなる韻と開音節単母音の韻[5]における母音をプロットしたものである（前者は点線、後者は実線）。点線と実線で同じ色の韻は、母音説において音韻的に同じ母音とされているもの同士（中舌母音については対応する長短のペア）であることを指す。図の楕円部に注目すると、「短母音＋舌背音」からなる韻（点線）のうち中・大の系列は同じ色の開音節単母音の韻（実線）に比べ開口度がやや大きくなっているほか、母音の初頭部分の音色がやや重なっていることがわかる。

[5] ただし、短母音は開音節に現れないので、ここでは 9 つの長母音のみをプロットしている。

このことからまず、ハノイ方言の「短母音＋舌背音」からなる各韻は以下のように音声表記すべきであることがわかる。

表 6: ハノイ方言の「短母音＋舌背音」からなる韻の新たな音声表記

開口度	短母音＋硬口蓋音の韻	短母音＋軟口蓋音の韻	短母音＋両唇軟口蓋音の韻
小	[i̯ɲ:], [i̯c:]	[ɪŋ:], [ɪk:]	[ɪ̯^{w}$\widehat{ŋm}$:], [ɪ̯^{w}$\widehat{kp}$:]
中	[ɜ̆^{j}ɲ:], [ɜ̆^{j}c:]	[ɜ̆ŋ:], [ɜ̆k:]	[ɜ̆^{w}$\widehat{ŋm}$:], [ɜ̆^{w}$\widehat{kp}$:]
大	[ă^{j}ɲ:], [ă^{j}c:]	[ăŋ:], [ăk:]	[ă^{w}$\widehat{ŋm}$:], [ă^{w}$\widehat{kp}$:]

5. 考察

本節では、前節のフォルマント分析の結果を踏まえて、二種類の分析方法を比較する。

5.1. 母音説

前節の表 6 の音声表記から、「短母音＋舌背音」からなる韻のうち中・大の系列の開口度は通常の長母音よりも大きいことがわかる。このことから、母音説の分析の派生規則 (1) は少なくとも以下のように修正されなければならない。

(4) 「短母音＋舌背音」からなる韻の母音説における派生規則（修正版）

	iŋ	eŋ	ɛŋ	uŋ	oŋ	ɔŋ
短母音化	ĭŋ	ĕŋ	ɛ̆ŋ	ŭŋ	ŏŋ	ɔ̆ŋ
口蓋化・唇音化	ĭɲ	ĕɲ	ɛ̆ɲ	ŭ$\widehat{ŋm}$	ŏ$\widehat{ŋm}$	ɔ̆$\widehat{ŋm}$
母音の異化	ɨ̆ɲ	ə̆ɲ	ɜ̆ɲ	ɨ̆$\widehat{ŋm}$	ə̆$\widehat{ŋm}$	ɜ̆$\widehat{ŋm}$
わたりの付加	ɨ̯^{j}ɲ	ə̯^{j}ɲ	ɜ̯^{j}ɲ	ɨ^{w}$\widehat{ŋm}$	ə^{w}$\widehat{ŋm}$	ɜ^{w}$\widehat{ŋm}$
<u>開口度の増大</u>	ɨ̯ɲ	ɜ̯^{j}ɲ	ă^{j}ɲ	ɨ^{w}$\widehat{ŋm}$	ɜ^{w}$\widehat{ŋm}$	ă^{w}$\widehat{ŋm}$

ここで新たに規則として加わった「開口度の増大」は、一連の韻のうち中・大の系列にのみ適用され、小の系列には適用されない。しかし、開口度が中・大の系列のみが特定の自然類をなすとは考えづらく、この新たな規則はやや不自然な規則となる。

またこの分析は、表層形の母音部分の音色についても問題がみられる。「短母音＋舌背音」からなる韻のうち、開口度が中の系列の母音は末子音に関わらずすべて [ɜ̆] であり、また開口度が大の系列の母音はすべて [ă] となっている。つまり、「短母音＋舌背音」からなる韻の対立を母音音素の対立と捉えると、各母音音素の異音の分布が以下のように互いに重なってしまうこととなる。

(5) 従来の分析における、母音音素の異音の分布（太字部分が異音の分布の重なり）

/e/	→ **[ɜ̆]** / _ [velar]	/ɜ̆/	→ **[ɜ̆]**	/o/	→ **[ɜ̆]** / _ [velar]
	→ [e] / elsewhere				→ [o] / elsewhere
/ɛ/	→ **[ă]** / _ [velar]	/ă/	→ **[ă]**	/ɔ/	→ **[ă]** / _ [velar]
	→ [ɛ] / elsewhere				→ [ɔ] / elsewhere

よって音響的な観点からみて、「短母音＋舌背音」からなる韻の対立を母音の違いとしてみるような母音説はやや不自然である。

5.2. 末子音説

前節において指摘した、「短母音＋舌背音」からなる韻の分析に関するふたつの問題は、当該の韻の対立を以下のように母音の違いではなく末子音の違いであると捉えると、うまく回避することができる。

表 7：本稿の主張する北部方言の「短母音＋舌背音」の韻の記述と分析（再掲）

/iŋ/, /ic/ → [i̯ŋ:], [i̯ʲc:]	/iŋ/, /ik/ → [iŋ:], [ik:]	/iŋ͡m/, /ik͡p/ → [i̯ʷŋ͡m:], [i̯ʷk͡p:]
/ɜŋ/, /ɜc/ → [ɜ̆ʲŋ:], [ɜ̆ʲc:]	/ɜŋ/, /ɜk/ → [ɜŋ:], [ɜk:]	/ɜŋ͡m/, /ɜk͡p/ → [ɜ̆ʷŋ͡m:], [ɜ̆ʷk͡p:]
/ăŋ/, /ăc/ → [ăʲŋ:], [ăʲc:]	/ăŋ/, /ăk/ → [ăŋ:], [ăk:]	/ăŋ͡m/, /ăk͡p/ → [ăʷŋ͡m:], [ăʷk͡p:]

上のように分析するとまず、「短母音＋舌背音」からなる韻の開口度は主母音 /i/, /ɜ/, /ă/ の開口度に依存していると説明することにより、前項の「開口度の増大」のような派生規則を組み込む必要がなくなるほか、規則の数自体が減る。また、(5) のような異音の重なりも解消される。よって音響的な観点から見ると、母音説より末子音説のほうが自然である。

(6) 本稿の主張する分析における派生規則

	iŋ	ɜŋ	ăŋ	iŋ͡m	ɜŋ͡m	ăŋ͡m
i の短母音化	ɨŋ	ɜŋ	ăŋ	ɨŋ͡m	ɜŋ͡m	ăŋ͡m
前舌化・後舌化	i̯ŋ	ɜŋ	ăŋ	i̯ŋ͡m	ɜŋ͡m	ăŋ͡m
わたりの付加	i̯ʲŋ	ɜ̆ʲŋ	ăʲŋ	i̯ʷŋ͡m	ɜ̆ʷŋ͡m	ăʷŋ͡m

6. おわりに

本稿では音響的観点から、ベトナム語北部方言の「短母音＋舌背音」からなる韻の分析として、末子音に音韻的対立を認め、母音は同一音素とみなす分析が妥当であることを主張した。ただし知覚の観点からも同様のことが言えるのかはまだ定かでなく、今後は当該の韻の弁別のキューが母音にあるのか末子音にあるのかを検討していく必要がある。

参考文献

Boersma, Paul & Weenink, David (2015) Praat: doing phonetics by computer [Computer program]. Version 6.0.08, retrieved 10 December 2015 from http://www.praat.org/

Cao Xuân Hạo (2007) *Tiếng Việt mấy vấn đề về ngữ âm-ngữ pháp-ngữ nghĩa*. Tái bản lần 3. Hà Nội: Nxb. Giáo dục.

Đoàn Thiện Thuật (2007) *Ngữ âm tiếng Việt*. Tái bản lần thứ 5. Hà Nội: Nxb. Đại học Quốc gia Hà Nội.

Nguyễn Phan Cảnh (1964) Vài ý kiến về vấn đề giải thuyết các phụ âm cuối trong tiếng Việt hiện đại. *Thông báo khoa học Văn học-Ngôn ngữ học*: 1964-1965, 2. Hà Nội: Đại học Tổng hợp Hà Nội.

清水政明（2007）『日・越文化理解のための双方向語学教材の開発のための研究』「新しいアジアとの交流」事業研究成果報告書．首都大学東京．

C5

まとまった文における中国語イントネーション：文タイプに基づいて

服部拓哉（大阪大学大学院）
u107964c@ecs.osaka-u.ac.jp

1　はじめに

本研究では，中国語北京官話話者による物語文と説明文の読み上げ音声を用いて，中国語イントネーションの発音実態を記述する。

従来のイントネーション研究は短文を対象にしたものが中心であり，まとまった文を自然に読み，話すためにはどのようなイントネーションを付ければよいのか，という観点からの分析は多くない。さらにまずいことに，声調言語である中国語においては，イントネーションの研究自体がきわめて少ない。現在でも，赵元任による分析の枠組みを超える研究はないようである（平井，2012）し，この枠組みもやはりまとまった文については考慮していない。

標準中国語の音声的な基礎となった北京官話は，4 種の語彙的声調をもち，ピンインと呼ばれるローマ字表記で母音の上に記号を付けて示す：第 1 声（高平板）「¯」，第 2 声（上昇）「´」，第 3 声（低平板）「ˇ」，第 4 声（下降）「`」。声調は，方言によって多様な調値動態が観察される。他方，イントネーションについては地域差はほとんど存在せず，使用声域や話速などの韻律的特徴を用いて，種々のモダリティや機能を表す（Chao, 1968）。

本研究では，短文の分析の枠組みがまとまった文にも当てはまるどうか，使用声域・話速について，文タイプによってどのような特徴があるのか，といった観点から検討を行う。

2　研究素材・研究方法

2.1. 読み上げ資料と参加者

本研究では，読み上げ資料として，物語文に文芸作品を，説明文にニュース記事を選定し，それぞれ地の文と発話文を含む 4 文を取り上げて使用する。その際，中国語文に音声記号と和訳を併記した。音声記号は IPA を使用した音素表記であり，Duanmu（2007）を参照した。声調を IPA で音素表記する場合は通常，第 1 声から第 4 声まで，それぞれ/ˊ/, /ˇ/, /ˋ/, /ˆ/が使用されるが，たとえば IPA の第 1 声がピンインの第 2 声になるなど紛らわしいので，本研究ではピンインの声調記号を利用した。原文を語ごとに分け，その下に音声記号を併記した。それぞれの談話的なまとまりの末尾で，かっこ内にシャープ記号と数字を記すことで通し番号を付けたが，これは考察で使用する。和訳は，物語文については既存の訳を，説明文については拙訳を使用した。詳細は次項以降を参照されたい。

読み上げ実験には，近畿圏在住で，北京官話を母方言とする男女各 2 名，計 4 名が参加した（Table 1）。赵説では地域差はないということであったが，Xu（1999）など地域差を考慮した論考もあり，また実際に官話方言の下位区分に属する言語変種同士であっても，相互に意思疎通できないほどかけ離れている場合もあるので，今回は対象を北京官話に統一した。その際，言語地図（张，2012）を用いて方言分布を確認した。

Table 1　参加者

	生育地	年代	性別
F1	北京市豊台区	20 代	女性
F2	河北省承徳市		
M1	南蒙古自治区シリンゴル盟		男性
M2	河北省廊坊市		

2.1.1. 物語文

鲁迅（1881–1936）の筆になる《故乡》（故郷）を使用する。この作品を選定したのは，日本だけでなく中国でも長い間中学生用の教科書に採用されており（萬，2016），中国文学の好例であると考えたためである。

本研究で分析対象として扱うのは，主人公の迅（シュン）が幼なじみの闰土（ルントウ）と再会する場面で，ここから 4 文（計 51 音節）を抜粋した。少年時代は兄弟のような仲だったが，厳しい身分社会の中で関係が変化してしまったことが描写されており，本作品におけるもっとも印象的な場面の 1 つである。なお，和訳末尾の「と背中に隠れている子どもを引き出した。」の部分は，本研究の対象範囲の直後から続く “便拖出躲在背后的孩子来，” の和訳である。文章的に不可分であるため，そのまま残した。

“老爷！……”(#1)
lǎuje
我 似乎 打了 一个 寒噤；(#2) 我 就 知道，(#3) 我们 之间
wǒː sɻ̩ːxʷuː tǎːlɤ́ íːkɤ́ xǽntɕìn wǒː tɕàu tʂɻ̩ːtàu wǒːmən tʂɻ̩ːtɕǣn
已经 隔了 一 层 可悲 的 厚 障壁了。(#4)
ǐːtɕəŋ kɤ́ːlɤ́ īː tsʰə́ŋ kʰɤ̌ːpāi tɤ́ xàu tʂàŋpʲìːlɤ́
我 也 说不出 话。(#5)
wǒː jěː ʂʷōːpʷùtʂʰʷūː xʷàː
他 回过头去 说，(#6) “水生，(#7) 给 老爷 磕头。” (#8)
tʰāː xʷəikʷòːtʰə́utɕʰʷỳː ʂʷōː ʂʷə̌iʂəŋ kə̌i lǎuje kʰɤ̌ːtʰə́u

和訳（井上，1932）

「旦那様」

と 1 つハッキリ言った。私はぞっとして身震いが出そうになった。なるほど私どもの間にはもはや悲しむべき隔てができたのかと思うと，私はもう話もできない。彼は頭を後ろに向け

「水生や，旦那様にお辞儀をしなさい」

と背中に隠れている子どもを引き出した。

2.1.2. 説明文

説明文では，《吴永宁之死：寒门孝子的极限迷途》（《人民网》2017.12.16）より事件の概要がわかる部分を 4 文（計 110 音節）抜粋し，分析対象として使用する。《人民网》は，《人民日报》のウェブ版である。《人民日报》は，中国共産党中央委員会の機関紙であり，ユネスコによって世界トップ 10 の新聞の 1 つに選ばれている，有力紙である。

11 月 8 日，(#9) 吴 永宁 从 华远・华中心 的 62
ʂɻ́ːīːɥèː pāːzɻ̩ː úː jʷŭŋnʲə́ŋ tsʰʷúŋ xʷáːɥæ̆nxʷáːtʂʷūŋɕīn tɤ́ lʲə̀uʂɻːàr
层 楼顶 附属物 的 平台 坠下。(#10)
tsʰə́ŋ lə́utʲə̌ŋ fʷùːʂʷŭːùː tɤ́ pʰʲə́ŋtʰái tʂʷə̀iɕàː
现场 的 血迹 表明，(#11) 吴 永宁 没有 当场 死亡，(#12)
ɕæ̀ntʂʰə̌ŋ tɤ́ ɕʷèːtɕìː pʲə̆umʲə́ŋ úː jʷŭŋnʲə́ŋ máijə̆u tāŋtʂʰə̌ŋ sɻ̌ːwáŋ
他 还 爬行了 三十 多 米 的 距离。(#13)
tʰāː xái pʰáːɕə́ŋlɤ́ sæ̀nʂɻ́ː tʷōː mʲǐː tɤ́ tɕʷỳːlʲíː
“我 喜欢 高空，(#14) 喜欢 爬，(#15) 喜欢 刺激。” (#16)
wǒː ɕǐːxʷæn kāukʰʷūŋ ɕǐːxʷæn pʰáː ɕǐːxʷæn tsʰɻ̀ːtɕīː
在 微博上，(#17) 吴 永宁 称 自己 是 国内 极限 高空
tsài wəipʷóːʂaŋ úː jʷŭŋnʲə́ŋ tʂʰə̄ŋ tsɻ̀ːtɕǐː ʂɻ̀ː kʷóːnèi tɕíːɕæ̀n kāukʰʷūŋ

运动　挑战　第一人，(#18)　目标　是　无　任何　保护　挑战
ɥìntʷùŋ　tʰʲǎuʈʂæ̀n　tʲì:ī:zʐ́n　mʷù:pʲâu　ʂʐ̩̀:　ú:　zʐ̀nxɤ̌:　pǎuxʷù:　tʰʲǎuʈʂæ̀n
全世界　的　高楼　大厦。(#19)
tɕʰʷǽnʂʐ̩̀:tɕè:　tɤ　kāuléu　tà:ʂà:

和訳（拙訳）
11 月 8 日，呉永寧は 62 階建ての華遠・華センターの屋上の縁から落下した。
現場の血は，呉永寧が即死せず，30m 以上這ったことを物語っている。
「高いところが好き，登ることが好き，刺激が好き。」
微博で，呉永寧は国内のエクストリームスポーツにおける先駆者を自称し，保護具を付けずに世界中の高層ビルに挑戦することが目標だとつぶやいていた。

2.2. 手順
実験は大阪大学無響室で行った。読み上げ資料はすべて簡体字で提示した。話速やポーズが不自然になるのを防ぐため，事前に十分な練習を行った。読み上げ時にはできるだけ感情を込めてもらった。途中で言い間違えやつまづきがあった場合は，第 1 文まで戻り，その文頭から再度読み上げてもらった。録音機材には TEAC 社の DR-40 と DR-05（サンプリング周波数：44.1kHz，量子化ビット数：16bit）を使用した。録音音声は音響分析用ソフト Praat（version 6.0.29）で測定した。
分析の際，話速は音節数を実発話時間（読み上げ時間とポーズ時間の差）で割ることによって算出した（音節／秒）。ポーズは，Praat の音声波形と，聴覚的判断を併用して確認した。ただし，境界の無声区間やきしみ声の時間はポーズに含めなかった。なお結果について述べるときは，煩瑣を防ぐため数値は小数第 2 位を四捨五入した。
きしみ声は，閉じた声門の前部から漏れて出る声で，ピッチ周期が不規則・音の強さが減少している・基本周波数（F_0）が低い，という特徴をもつ。標準中国語においては第 3，4 声と高い相関がある（Belotel-Grenié and Grenié, 1994, 1995）。本研究の範囲でも，声帯振動が不安定であるため F_0 が記録されなかった部分が多く，また記録できた部分も正確に計測できているかどうかわからない。一方で，イントネーションとして特別な語用論的な意味をもっているというわけでもなく（Chao, 1968），文や段落，韻律グループの末尾において境界表示を行うという機能しかもっていない（Belotel-Grenié & Grenié, 2004）ため，ピッチに限り，本研究では一貫してきしみ声の部分を分析の対象外とする。

3　結果・考察
3.1. 短文の記述の検証
Chao（1968）による分析項目の中で，本研究の対象範囲と合致するのは，(1)長めの文における漸降性，(2)未完の部分における高音域，(3)命令表現の末尾数音節における話速の上昇の 3 項目である。以下，順を追って検討を行う。
(1)長めの文における漸降性：何語においても，文末に近づくにつれ，ピッチが徐々に低くなっていく傾向があり，これは漸降性と呼ばれている。中国語においても，長め（5 音節以上）の文が，特別な語用論的意味を伴わなずに発される場合，漸降性の傾向が現れる。
以下，各話者について，まとまりごとの声域の平均値をまとめた（Tables 2, 3）。数値は，各まとまり内におけるピッチの最高値と最低値の平均（Müller, 2007）で，単位は半音（基準値は 100Hz）である。表中の各太枠内で，上から下に向かって文末に近づくということになるが，漸降性があるのであれば数値が徐々に低くなっていくはずである。第 1 文（#1）は，5 音節未満なので，下表から除外した。第 3 文は単文なので，平均値は文頭の “我”（上枠；第 2 声に声調変化）と文末の “话”（下枠）について算出した。ただし，F1 のみ “话” をきしみ声で発声していたので，その直前の “说不出” の平均値を記録した（同様に下枠）。

Table 2　声域の平均値（物語文）

		F1	F2	M1	M2
第２文	#2	14.4	12.8	4.5	4.3
	#3	12.5	15.4	5.0	5.5
	#4	14.2	12.4	5.9	4.7
第３文	#5	16.5	12.2	2.0	3.2
		10.9	14.2	6.0	2.2
第４文	#6	13.4	15.9	6.4	2.8
	#7	14.6	15.5	3.6	5.2
	#8	13.9	12.5	2.8	3.4

Table 3　声域の平均値（説明文）

		F1	F2	M1	M2
第１文	#9	15.0	13.9	3.9	9.1
	#10	13.9	13.3	4.9	4.8
第２文	#11	14.3	10.3	6.2	7.0
	#12	14.8	12.6	3.8	3.8
	#13	13.6	13.3	4.6	2.8
第３文	#14	17.6	15.9	5.0	5.9
	#15	9.9	11.7	4.1	2.6
	#16	13.3	13.9	7.0	5.6
第４文	#17	12.5	14.7	4.7	7.9
	#18	13.5	12.2	5.0	3.8
	#19	14.7	14.1	6.3	5.3

最後のまとまりが最初のまとまりより少しでもピッチが低かったのは，対象の 28 例中 17 例で 6 割程度（60.7%）であった。内訳として，物語文では 12 例中 6 例で 5 割，説明文では 16 例中 11 例で 7 割程度（68.8%）と，説明文のほうでより漸降性の影響が出ていた。これは説明文の音節数が相対的に多いことも一因のように思えるが，次の(2)の結果を考慮すると，そうとも言い切れない。これらは文タイプとはとくに関連はないように思える。

(2)未完の部分における高音域：未完の，つまり文中の句・節は，文末の結びの句・節より全体が高く発される。英語において，発言の継続を表す場合に上昇調が用いられるのと似ているが，中国語におけるそれとは次の 2 点で異なる：(1)かすかな変化，(2)一部の曲線的な変化ではなく全体的な変化。

漸降性の分析の際には，最初と最後のまとまりのピッチの平均値を比較したが，ここでは文中（最後以外すべて）のまとまりと文末（最後）のまとまりのピッチの平均値を比較した（Tables 2, 3）。該当したのは，28 例中 10 例で，3～4 割程度（35.7%）であった。物語文では 12 例中 5 例で 4 割程度(41.7%),説明文では 16 例中 5 例で 3 割程度(31.3%)と，ここでは漸降性の結果とは異なり，説明文のほうが該当の割合が低かった。

(3)命令表現の末尾数音節における話速の上昇

特別な含意のない単純な命令表現では，文の末尾の数音節が，文末に近づくにつれて話速がかすかに上昇する。

本研究の対象範囲では, 物語文の第 4 文後半部, #8 の “给老爷磕头” が命令表現である。話者別に各音節の話速を示す（Table 4, Figure 1）。単位は音節／秒である。

Table 4　命令表現の話速

	给	老	爷	磕	头
F1	16.1	10.8	5.4	6.4	3.8
F2	6.8	5.7	7.9	5.2	3.0
M1	13.2	5.6	6.9	5.1	3.0
M2	16.9	6.6	13.5	6.5	4.7

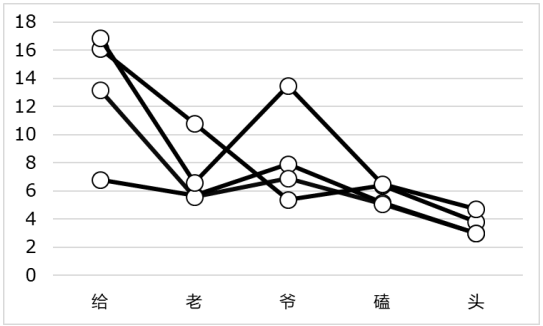


Figure 1　命令表現の話速

この図表を見ると，命令表現の末尾数音節で話速が上昇するということはないとわかる。

どちらかと言えば，どの話者もむしろ末尾のほうが相対的に話速が低い，つまり遅いほうに落ち着いていっている。

以上の検討より，短文の記述の中で本データの過半数について言えるのは(1)の漸降性に関する記述のみで，ほかの項目についてはあまり当てはまらないということがわかった。ただし，本研究で扱ったのは Chao（1968）による 10 の分析項目のうち 3 つだけである。このように部分的な検証ではあるが，それでも短文の記述を盲目的に，まとまった文に当てはめるのは危険だということがわかる。

3.2. 使用声域

本研究の分析範囲全体について，文タイプごとのピッチ変動を話者別に示した（Figure 2）。縦軸は高さ（半音；どの話者の音域も 24 半音分で統一），横軸は時間（秒）を示す。

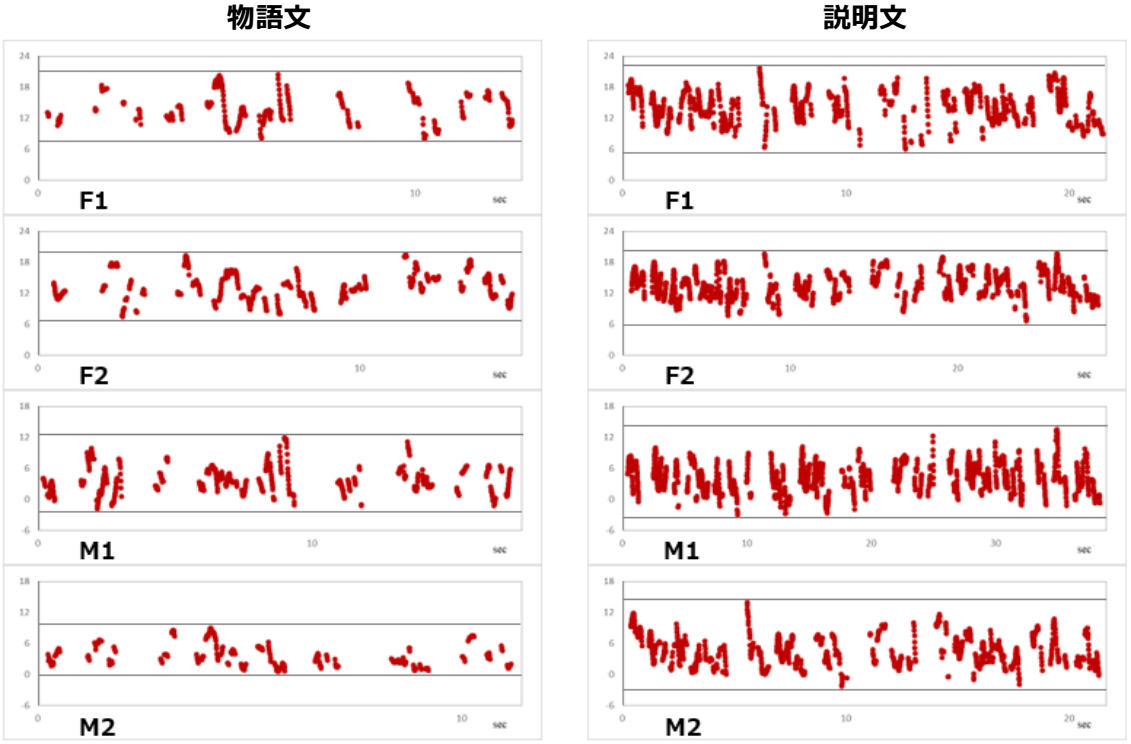


Figure 2　使用声域

上の図をまず縦で見る，すなわち同じ文タイプを話者ごとに見てみると，使用声域が相対的に広い話者と狭い話者がいることが確認できる。使用声域の広さは，個人差が大きいということがわかる。

次に，上の図を横で見る，すなわち同じ話者を文タイプごとに見てみると，物語文と説明文で比較したとき，その広さの差は話者によって大きく異なるが，どの話者においても相対的に説明文のほうが，ピッチ変動幅が広いことがわかる。たとえば英語では，会話（Denes and Pinson, 1993）と比較して，客観性が要求される説明文ではピッチ変動幅が 1 オクターブ抑えられる（Lihiste, 1970）が，中国語ではこの限りではないようである。物語文・説明文の各文タイプの地の文同士・発話文同士を比較しても同様に，むしろ地の文のほうがピッチ変動幅が広かった。中国語では主観性・客観性はあまりピッチ変動幅に関係しない，あるいは客観的な文のピッチ変動幅は逆に広くなるという可能性がある。文の長さ，すなわち音節数に影響を受けている可能性も排除しきれないが，この点も含め，今後統合的に検討したい。

3.3. 話速

以下, 話者別に文タイプごとの話速を示す（Table 5, Figure 3）。単位は音節／秒である。

Table 5　文タイプごとの話速

	物語文	説明文
F1	5.9	6.0
F2	5.3	4.9
M1	5.0	4.2
M2	7.2	6.4

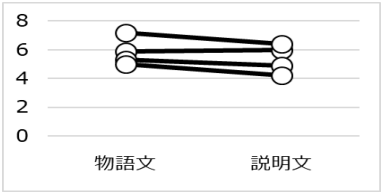


Figure 3　文タイプごとの話速

上の図表を見ると，説明文のほうが遅い傾向にあるが，その差はどの話者も 1 音節／秒以下とわずかなものであり，また逆に説明文のほうを速く発している話者（F1）もいるように，すべての話者に当てはまるわけでもないので，取り立てて論ずるほどのことはないように思える。

以上の検討より，文タイプによって変化する中国語イントネーションの主な領域は，話速ではなく使用声域であることがわかった。

4　結論

中国語北京官話話者による，物語文と説明文の読み上げ音声を用いて，中国語イントネーションの発音実態を分析した結果の概略を以下にまとめる：

(1) 短文の分析の枠組みで，まとまった文についても言えそうのは，漸降性に関する記述ぐらいであった。
(2) 使用声域は，たとえば英語とは異なり物語文より説明文のほうが広かった。
(3) 話速については，物語文と説明文でとくに違いは見られなかった。

References

Belotel-Grenié, A., and Grenié, M. (1994) "Phonation types analysis in Standard Chinese", *Proceedings of ICSLP'94,* 343–346.

Belotel-Grenié, A., and Grenié, M. (1995) "Consonants and vowels influence on phonation types in isolated words in Standard Chinese", *Proceedings of XIIIth ICPhS,* 400–403.

Belotel-Grenié, A., and Grenié, M. (2004) "The creaky voice phonation and the organisation of chinese discourse", *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages,* 5–8.

Chao, Y. R. (1968) *A grammar of spoken Chinese.* Berkeley: University of California Press.

Denes, P. B., and Pinson, E. N. (1993) *The speech chain: The physics and biology of spoken language.* Long Grove: Waveland Press.

Duanmu, S. (2007) *The phonology of Standard Chinese, 2nd edition.* Oxford: Oxford University Press.

Lehiste, I. (1970) *Suprasegmentals.* Cambridge: The MIT Press.

Müller, C. (2007) *Speaker classification I: Fundamentals, features, and methods.* Berlin: Springer.

Xu, Y. (1999) "Effects of tone and focus on the formation and alignment of f0 contours", *Journal of Phonetics* 27, 55–105.

井上紅梅（1932）『魯迅全集』東京：改造社

平井勝利（2012）『教師のための中国語音声学』東京：白帝社

萬確（2016）「日中の教科書における『故郷』（魯迅）の学習課題の比較：『作者の意図』と『解釈・鑑賞』」『岩大語文』(21), 117–128.

张振兴．(2012)『中国语言地图集：汉语方言卷』北京：商务印书馆

C6

韓国語ソウル方言における語中閉鎖音の知覚

邊 姫京（国際教養大学）
byun@aiu.ac.jp

1. はじめに

邊 (2017) はソウル方言における語中閉鎖音の音響特徴を明らかにするために関連が指摘されているパラメータのうち，VOT，閉鎖区間長（CD)，全長（TD＝VOT＋CD)，先行母音長，後続母音の fo について検討した．それによれば，いずれのパラメータも単独では 3 種の閉鎖音を区別することができないが，VOT，CD，TD は 3 種の閉鎖音を 2 つのグループに区別することができる．そして VOT と CD，あるいは VOT と TD のように 2 つのパラメータを組み合わせることで 3 種の閉鎖音をそれぞれのカテゴリーに分けることができる．具体的には，図 1 で見るように，激音とそれ以外の子音（濃音・平音）は VOT の違いで，濃音と平音は CD または TD の違いで分けられる．つまり，激音は長い VOT，濃音は短い VOT と長い CD（または TD)，平音は短い VOT と短い CD（または TD）に特徴づけられる．

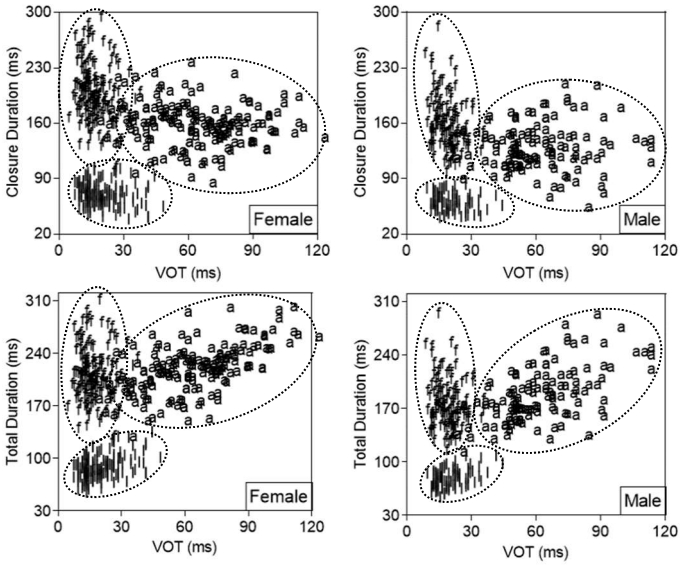


図 1: 語中閉鎖音の生成におけるカテゴリー域（邊（2017）の図 8）
横軸：VOT，縦軸：CD（上段）と TD（下段），図中の a：激音/tʰ/ /kʰ/，l：平音：/t/ /k/，f：濃音/t*//k*/

本研究は，生成における音響特徴が知覚においても有効であるかを調べたものである．VOT と CD に焦点を当て，2 つのパラメータを段階的に変化させた合成音を作成し，ソウル及び京畿道出身の韓国語母語話者を対象に聴取実験を行った．以下にその詳細を報告する．

2. 研究方法

2.1. 刺激音

原音は，1987 年生まれのソウル方言話者（男性）が発話した濃音の「ak*a（아까）」である．原音をどの子音にするかは先行研究でもまちまちであるが，Oh（2019）は，激音と平音の場合，どちらを使っても結果には影響しないことを報告している．本研究では元の音色を変えずに操作できたのが濃音だったので濃音を使用した．VOT の操作には Praat の manipulate 機能を利用したが，使える音声のうち激音は圧縮によりバーストの部分が強くなり過ぎて平音の音色が出にくく，平音は濃音と激音に感じられる緊張が感じられないので使用しなかった．軟口蓋音にしたのは，VOT 操作のためにある程度の長さの VOT が必要だったからである．

VOT と CD の幅は，邊（2017）の結果を基に，それぞれ 5～85ms，40～240ms に決めた．図 2 は 1990 年代生まれの男性（M）と女性（F）の発話結果である．図 1 は 1953～1999 年生まれの話者の結果であるが，今回の聴取実験に参加する参加者を 1990 年代以降生まれの若者に限定したので刺激音の幅も 1990 年代生まれの話者の発話を参照した．原音の VOT は 25ms，CD は 137ms である．CD は 140ms にしてから操作を開始した．まずは VOT を 10ms 間隔で 9 段階に変化させ，その後 CD を 20ms 間隔で 11 段階に変化させた．

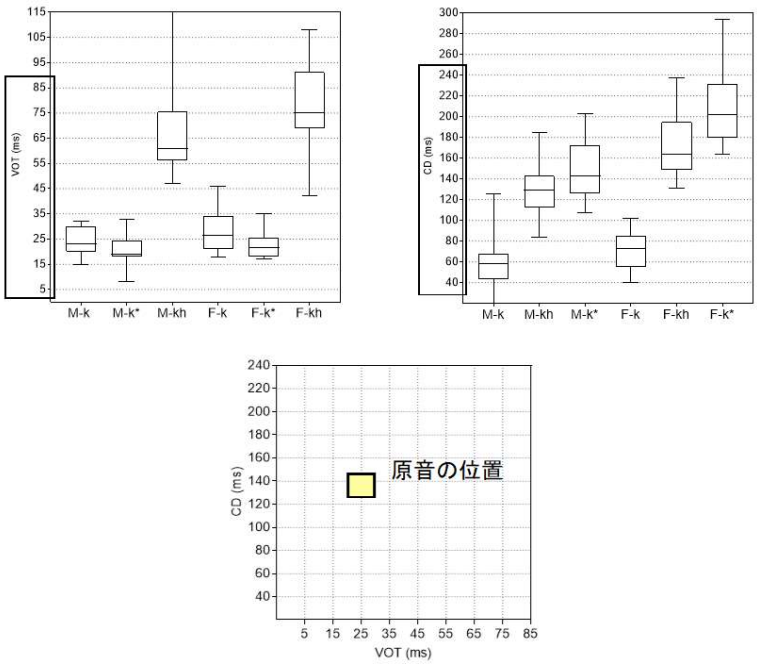


図 2: 軟口蓋音/ak*a/の測定値（1990 年代生まれのソウル方言話者，邊（2017）から作成）

2.2. 参加者（聴者）

ソウル及び京畿道で生まれ育った 1990～2001 年生まれの男女 34 名で，全員大学生である．

2.3. 手順

実験は大学の研究室で，Praat 上で行った．参加者はヘッドホーンから流れる音声を聴いてもっとも近いと思われるものをパソコン画面上に表示されるハングル表記の「aka」「ak^{h}a」「ak*a」からクリックして選ぶ．刺激音は必要であればさらに 2 回まで聴き直すことができる．回答を間違えた場合は再回答ができるように設定した．本実験に先立って 6 問を使って本実験と同じ形式で練習を行った．本実験は 100 問（刺激音 99 問とダミー1 問）で，25 問ごとに短い休憩がある．回答に制限時間は設けず参加者の裁量に任せたが，かかった時間は長い人でも 5 分以内であった．事前のアンケートで聴力に問題があると報告した参加者はいなかった．参加者には実験後に謝礼を渡した．

3. 結果

3.1. パラメータ別結果

図 3 に VOT，図 4 に CD の結果を示す．図 3 で見るように VOT が長くなるほど激音の同定率が上がり，逆に VOT が短いほど濃音あるいは平音の同定率が上がる．濃音と平音は 35ms 以上では両者の間に差がほとんどない．両者の同定率に差が出るのは 25ms 以下で，濃音の同定率は平音のそれより倍以上高い．激音と濃音は 30ms を境にきれいに分かれており，VOT が激音と濃音を区別する主要なパラメータであることがわかる．つまり，長い VOT であれば激音，短い VOT であれば濃音に優先的に同定されると言える．ただし，激音が VOT のみで 9 割近くまで同定されるのに対して，濃音はもっとも短い VOT でも同定率は 7 割程度で，VOT だけで濃音が同定できるとは限らない（図 3 では 15ms 以下で同定率にほとんど変化がない）．濃音と同様に短い VOT に反応する平音ももっとも短い VOT での同定率は 3 割程度で，VOT だけの同定は濃音よりもさらに精度が落ちる．

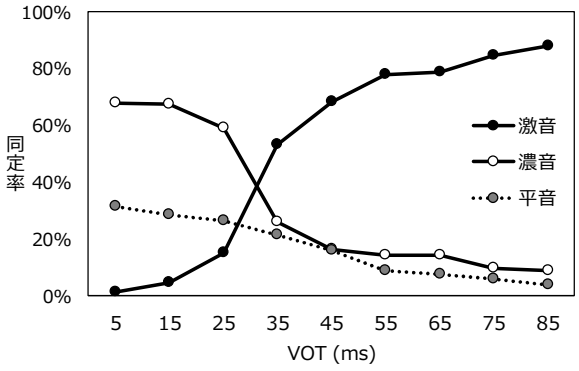


図 3: VOT の同定率

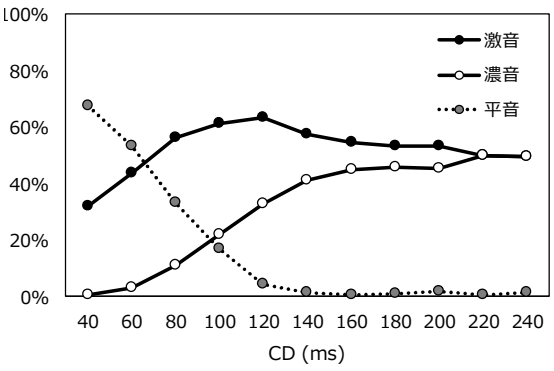


図 4: CD（閉鎖区間長）の同定率

図 4 の CD の結果を見よう．平音は 100ms 以下であれば，CD が短いほど同定率が上がる．図 4 でモットも短い値は 40ms であるが，40ms 時点でも平音のグラフは上向きになっており，さらに短い値であれば同定率はさらに上がることが予想される．この CD の結果と前述の VOT の結果を合わせると，平音の同定にかかわる主要なパラメータは CD であり，

VOT も手がかりになると言えそうである．濃音は CD が長いほど同定率が上がる傾向にあるが，これが言えるのはおよそ 160ms までで，それ以上は長くなっても大きな変化はない．激音は 80ms 以上では同定率にあまり差がなく，CD の効果は VOT に比べて低いと言えよう．

図 3，4 の結果を整理すると，激音の知覚には主に長い VOT が用いられる．濃音の知覚には短い VOT と長い CD が用いられる．平音の知覚には短い VOT と短い CD が用いられる．

3.2. VOT と CD の組み合わせの結果

図 5，6 に VOT と CD を組み合わせた 3 次元グラフを示す．横軸は VOT (ms)，縦軸は CD (ms)，各セルの濃さは同定率を表す．図 5 は各子音の同定率を 20%間隔で示した．図 6 は当該子音の同定率で，子音の記載のない 50%のセルは，隣接する子音と同定率が 50%で同値の場合である．

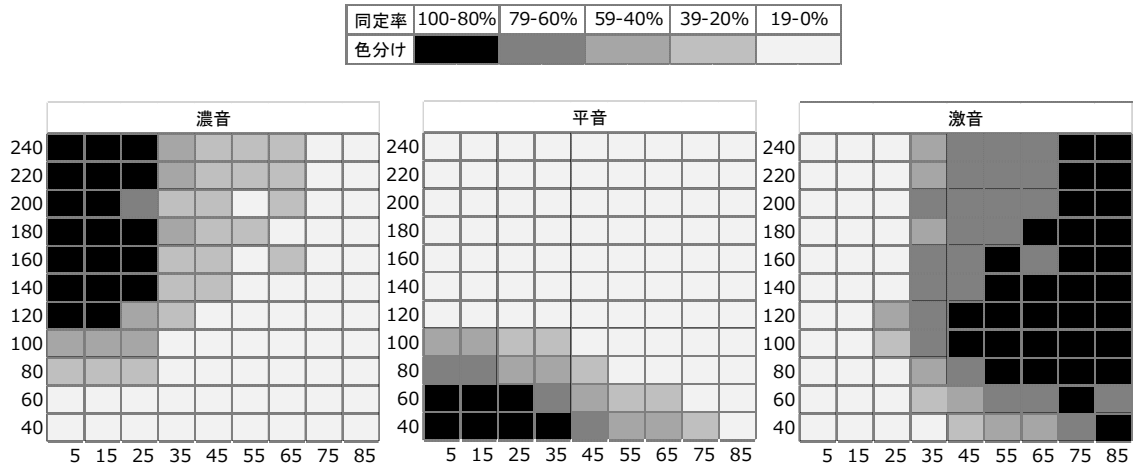


図 5: 各子音の同定率

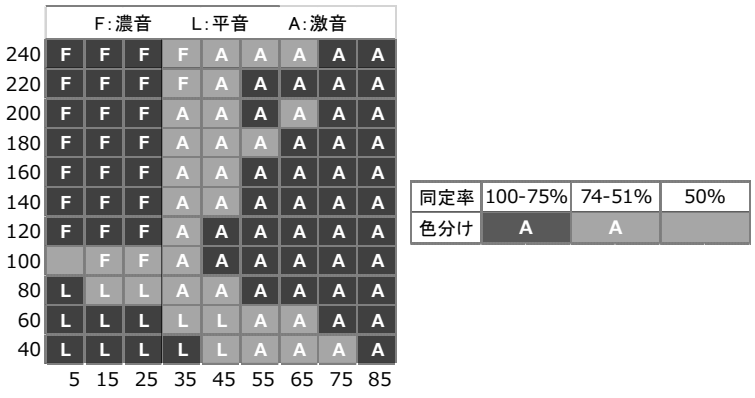


図 6: 3 種の閉鎖音の同定率（語中閉鎖音の知覚におけるカテゴリー域）

図 5 で見るように，濃音は短い VOT と長い CD，平音は短い VOT と短い CD，激音は概して 55ms 以上であれば CD と関係なく VOT のみで知覚されている．図 6 では各子音のカ

テゴリー域の相対的な位置関係を見ることができる．VOT の違いで激音とそれ以外の子音（濃音と平音），CD の違いで濃音と平音が区別されており，これは図 1 で見た生成におけるカテゴリー域と一致する．これにより語中の 3 種の閉鎖音は，生成，知覚のいずれも VOT と CD の違いで区別されると言える．

3.3. パラメータの影響度

同定における各パラメータの相対的な影響力を検討するために多項ロジスティック回帰分析を行った（RStudio 使用 Version 1.1.456）．分析には R（R Core Team 2016）のパッケージ nnet の multinom 関数を用いた．説明変数は VOT と CD，目的変数は子音種（激音，濃音，平音の 3 択）である．VOT と CD は中心化を行った．参照カテゴリーは激音である．

表 1 に回帰分析の結果，表 2 にオッズ比と信頼区間を示す．表 1 の係数 β（推定値）は，プラスの値であれば参照カテゴリーである激音に対して濃音または平音に同定される確率が高く，マイナスの値であれば参照カテゴリーである激音に同定される確率が高いことを表す．また，プラス，マイナスの符合を除いた値（絶対値）は目的変数への影響の大きさを表す．具体的には，濃音の CD の係数はプラスなので濃音に同定される確率が高く，濃音の VOT と平音の VOT，CD の係数はマイナスなので参照カテゴリーである激音に同定される確率が高い．係数の絶対値からは，激音と濃音，激音と平音の知覚には CD よりも VOT の影響力が強く（激音 vs 濃音：VOT の|-0.073|＞CD の|0.011|，激音 vs 平音：VOT の|-0.077|＞CD の|-0.047|），濃音と平音の知覚には VOT よりも CD の影響力が強いことがわかる（濃音 vs 平音：CD の|0.011-(-0.047)|＞VOT の|(-0.073)-(-0.077)|）．なお，VOT と CD の P 値はいずれも有意である．

表 1: 多項ロジスティック回帰分析の結果

	濃音（激音 vs 濃音）				平音（激音 vs 平音）			
	係数 β	標準誤差	z 値	P 値	係数 β	標準誤差	z 値	P 値
切片	-0.843	0.060	-14.141	<.0001	-3.381	0.174	-19.405	<.0001
VOT	-0.073	0.003	-27.611	<.0001	-0.077	0.004	-21.283	<.0001
CD	0.011	0.001	11.077	<.0001	-0.047	0.002	-19.175	<.0001

表 2: オッズ比と信頼区間

	濃音（激音 vs 濃音）			平音（激音 vs 平音）		
	オッズ比	2.5%	97.5%	オッズ比	2.5%	97.5%
切片	0.430	0.383	0.484	0.034	0.024	0.048
VOT	0.930	0.925	0.934	0.926	0.919	0.932
CD	1.011	1.009	1.013	0.955	0.950	0.959

では，具体的にどのくらい影響力が強いかは，表 2 のオッズ比（推定値）より推定することができる．オッズ比は 1 を基準に 1 より大きい場合は当該のカテゴリー，1 より小さい場合は参照カテゴリーに同定される確率を表す．つまり，濃音の CD のオッズ比は 1 より大きい 1.011 なので，CD が長いほど激音に対して濃音に同定される確率が 1.011 倍高い．また，オッズ比が 1 より小さい濃音の VOT と平音の VOT，CD は，VOT または CD が長いほど参照カテゴリーである激音に同定される確率がそれぞれ 0.93 倍，0.926 倍，0.955 倍高いことになる．オッズ比は，説明変数が連続変数の場合は 1 単位増すごとの増加分になるので，激音と濃音では CD が 1ms 増すごとに激音より濃音に同定される確率が約 1%増すと解釈できる．VOT は 1ms 増すごとに濃音より激音に同定される確率が約 8%増す（激音に対して濃音に同定される確率である 0.930 の逆数，1/0.93≒1.075）．同様に，VOT と CD が 1ms 増すごとに平音に対して激音に同定される確率はそれぞれ約 8%と 5%増す（VOT：1/0.926≒1.079，CD：1/0.955≒1.047）．濃音と平音の場合，VOT と CD が 1ms 増すごとに平音に対して濃音に同定される確率はそれぞれ約 0.4%と 6%になる（VOT：0.930/0.926≒1.004，CD：1.011/0.955≒1.058）．

4. まとめ

以上，生成における音響特徴が知覚においても有効であるかを，VOT と CD を対象に検討した．結果で明らかになったように，VOT と CD は生成と同様に知覚においても有効なパラメータであり，カテゴリー間の相対的な位置関係も生成の場合とほぼ同様である．以下に結果をまとめる．

語中における 3 種の閉鎖音は，VOT と CD の組み合わせで，長い VOT であれば激音，短い VOT と長い CD であれば濃音，短い VOT と短い CD であれば平音と知覚される．激音と濃音，激音と平音における VOT の境界はともに 30ms 前後，激音と平音，濃音と平音における CD の境界はそれぞれ 60ms 前後，100ms 前後であった．この境界は軟口蓋音の場合で，軟口蓋音より VOT が短い両唇音，歯茎音では異なり得る．多項ロジスティック回帰分析では係数 β の比較とオッズ比の比較から各パラメータの相対的な影響力について議論した．

本研究は、JSPS KAKENHI 17K02685 の助成を受けたものである．

参考文献

邊姫京（2017）「韓国語ソウル方言における語中閉鎖音の音響特徴」『音声研究』 21:2, 61-79.

Oh, Eunjin. (2019) "The effect of base token selection for stimuli manipulation on the perception of lenis and aspirated stops in Seoul Korean. " *Proceedings of the 2019 Spring Conference of Korean Society of Speech Sciences*, 79.

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

C7

朝鮮語ソウル方言における語頭破裂音の新しい音響パラメータの提案[1]

山崎 亜希子（早稲田大学）

1. はじめに

本発表では，朝鮮語ソウル方言（以下，ソウル方言）の破裂音を対象に，語頭における 3 系列子音（平音・激音・濃音）の対立を支える音響特徴として，子音区間の高周波数帯域の噪音成分の強度（パワー）を提案する．

1.1. 平音と激音の対立を支える音響特徴：先行研究

朝鮮語の破裂音には，平音/p, t, k/，激音/pʰ, tʰ, kʰ/，濃音/p’, t’, k’/という 3 系列の対立がある．これらは，語頭の位置ではすべて無声音で実現することから，これらの対立を支える音響特徴は伝統的に「激音＞平音＞濃音」の順に長い VOT とされてきた．ところが，2000 年代に入り，ソウル方言では平音と激音の VOT が重複していると主張されるようになる（Silva 2006 など）. なお, 濃音は従来どおり, VOT が最も短いという認識で共通している.

そこで，注目されるようになったのが，語頭子音に後続する母音の F0 である．語頭子音が平音ならば，第 1 音節と第 2 音節の高さが「LH」，激音・濃音・摩擦音であれば「HH」で現れることから，平音と激音において差がなくなった VOT に代わり，この F0 が平音と激音の弁別特徴になったと主張されるようになり，現在，これが広く受け入れられている．

ところが，F0 について被験者データを平均すると，たしかに平音とそれ以外（激音・濃音）とでは分布が重ならず，明瞭に異なっているが，被験者個別のデータを観察すると，「平音」と「激音・濃音」の分布の差が小さく，さらには重なる被験者もいる．そのような被験者にとっては，ほかの被験者に比べて，F0 が平音と激音の対立を保つ音響特徴になりにくいと考えられる．そこで，本論文では新たに，平音と激音の対立を支える音響特徴として子音（VOT）区間の噪音成分の観察を導入し，その特徴が平音と激音の区別に有効に働く可能性を示す.

1.2. 「高周波数帯域の噪音成分のパワー」とは何か？

「パワースペクトル」とは，ある音の区間にどのような周波数成分が含まれているのかを示したもので，その区間を構成する周波数成分のどこが強いか，弱いかがわかる（坂本 真一 ほか 2016）．山崎亜希子（2014）では，本発表とは異なり，語中の分析ではあるものの，パワースペクトルの観察を通じて，平音と激音では 4000Hz-6000Hz 付近の噪音成分のパワーの大きさに違いがあることを明らかにした．しかし，その噪音成分のパワー違いが子音区間内で一時的（たとえば，子音開始部分のみ）なのか，持続しているものなのか，この方法ではパワーの時間的変化を動的に捉えることはできなかった：

[1] 本発表は, 東京外国語大学に提出した博士論文の成果の一部に加筆・再構成したものである．また，本発表の一部は，JSPS 科研費 JP19K00597 の助成を受けたものです.

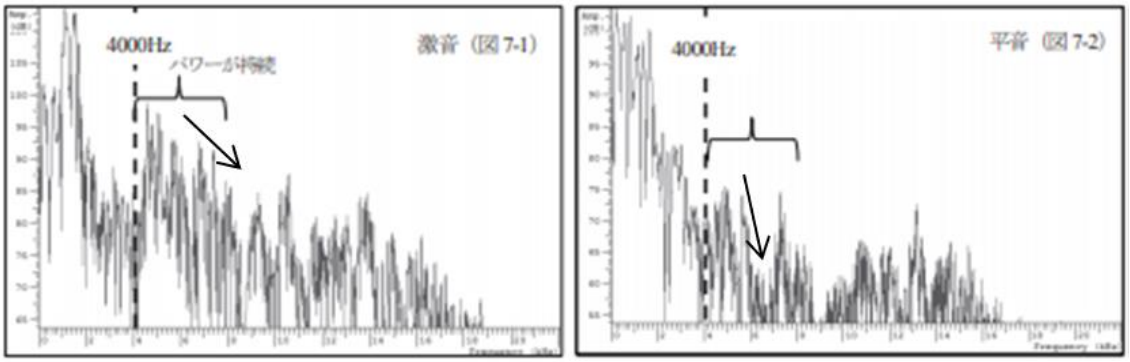


図 1: VOT 区間のパワースペクトル（山崎 2014: 129 図 7 引用. 左:激音, 右:平音） ※矢印は発表者追加

そこで本発表では，高周波数帯域の噪音成分のパワーが子音区間内でどう変化していくのかを動的に捉えるために，子音区間を 10ms ごとに区切り，パワーの推移を観察した結果を示す．ここでの「高周波数帯域の噪音成分のパワー」とは「6000Hz から 7000Hz までの帯域の噪音成分の強さ（単位：dB)」のことである．この高周波数帯域は通常，摩擦音の観察で使用する帯域である．調音位置の違いから起こる噪音成分の違いを観察する指標と考えられ，摩擦音どうしを弁別する特徴が現れるとされる．たとえば「/ʃ/の同定は 2500Hz のピークに関連しているのに対し, /s/の同定は 5000-8000Hz のエネルギーピークに依存している」（Kent and Read 1992: 124）という．本発表では破裂音においても，この高周波数帯の噪音成分の違いが平音と激音の対立を支える音響特徴になると考え，音響パラメータとして採用した．つまり，破裂音の子音（VOT）区間の噪音成分を観察する．

2. 実験

2.1. 被験者

ソウル出身で 1982～86 年生まれの 4 名（女性 : FS1 氏，FS2 氏，男性 : FM1 氏，FM2 氏）を被験者（表 1）として発話実験を行った．すべての被験者の両親がソウル近郊（京畿道，仁川（インチョン）広域市）出身であることから，被験者は成長過程で家庭内でもソウル方言を使用してきたと推測できる：

表 1:被験者情報

被験者名	性別	生年	父親出身地	母親出身地
FS1	女性	1982 年	京畿道～13 歳ソウル	ソウル
FS2	女性	1983 年	京畿道～20 歳ソウル	京畿道～20 歳ソウル
MS1	男性	1982 年	仁川	ソウル
MS2	男性	1986 年	ソウル	仁川～20 歳ソウル

2.2. 実験語とキャリアセンテンス

実験語は，ターゲット子音（C）の/p, pʰ, p’/（以下，P 類），/t, tʰ, t’/（以下，T 類），/k, kʰ, k’/（以下，K 類）が語頭にあり，それに後続母音（V）/a/を組み合わせた開音節/CV/の 1 音節語（/pa, pʰa, p’a/，/ta, tʰa, t’a/，/ka, kʰa, k’a/）である．網羅的に組み合わせた 1 音節語であるが，実単語（たとえば，/pʰa/（파/ネギ）など）も存在する．

キャリアセンテンス「____가 아닌 것 같아요.（/___ka anin kɔt kathayo/）」（「____ではないと思います.」）の下線部分に実験語を入れたもの（以下，実験文）を発話してもらった．実験語（下線部分）に後続する/ka/は主格助詞であり，実験語とつながっていわゆる文節をなしている．この，実験語（1 音節）に主格助詞/ka/がついた文節単位を「単語」と呼ぶことにする．

発話録音作業は，東京外国語大学音声学実験室内の防音室とソウル市内にある録音スタジオで行った（共に，サンプリングレート 44.10kHz，16bit 量子化）．

録音は，実験語をランダムに配列した実験文リストを渡し，それを読み上げてもらう方式で行った．被験者に実験の意図は伝えず，発話時の注意点として，1）1 文読むごとに数秒間休止を入れ，複数の文を一息で読まないこと，2）実験文を途中で区切ったり強調したりせずに 1 文として自然な速度で読むように指示した．1 名につき，3 セット行った．1 名につき 27 データ（9 語×3 セット），計 108 データ（被験者 4 名分）を収集した．

2.3. 測定手順

破裂音の平音と激音（P 類/p, pʰ/，T 類/t, tʰ/，K 類/k, kʰ/）を対象に，語頭の子音区間における 6000-7000Hz 帯域の噪音成分のパワー変化を 10ms ごとに測定した．「子音区間」とは「VOT」（バースト時点からボイスバーの開始時点までの区間）と同じ区間である．濃音は，VOT の長さが 10ms 前後しかないため動的変化が観察しにくいため除外する．

音声の抽出・測定には，Praat（5.3.57，5.3.64，6.0.28，6.0.39）および Spit Editor（0.5.0.0）（佐藤大和・益子幸江 2013）を使用した．手順は次の通りである：

① Praat を使って，各実験語の VOT 区間のみの音声ファイル（wav 形式）を作成．
② ①のファイルを Praat の Filter（Pass Han Band）機能を使って，周波数帯の抽出範囲を 6000-70000KHz に指定し，抽出した音声ファイル（wav 形式）を作成．
③ ②の音声ファイルを Spit Editor に読み込み，10ms ごとにピッチマークを入れて，パワー算出し，グラフ化（縦軸：パワー（dB），横軸：VOT（ms））．

縦軸の「dB」は音圧の絶対値ではなく，相対的なレベルである．「何らかの値を基準」として周波数成分のレベルがそれより何 dB 低いか，高いか，というふうに，強弱を相対的に確認するものである（坂本真一 ほか 2016: 113）．本発表でも dB 値はパワーの相対的な大小の比較に用いる．10dB の差は 10 倍，20dB の差は 10^2 倍，つまり 100 倍となる．

3. 結果・考察

3.1. 子音区間における平音と激音の高周波数帯域の噪音成分のパワー比較:「平音＜激音」

図 2 は，平音/pa/（実線）と激音/pʰa/（点線）の子音区間の高周波数帯域のパワーを比較したグラフであり，各被験者につき各 3 回分の発話データすべてを示している．縦軸はパワー（強さ）（単位：dB），横軸は子音区間の長さ（単位：ms），つまり VOT 区間である．子音区間 0-10ms の値がグラフの横軸「10ms」にプロットされるため，「0ms」はグラフ上にはない．また，子音区間の最後の部分は 10ms より短い端数は，繰り上げてプロットしてある．たとえば，子音区間（VOT）が 87ms であった場合，80-87ms 区間のパワー値は 90ms 時点にプロットされている：

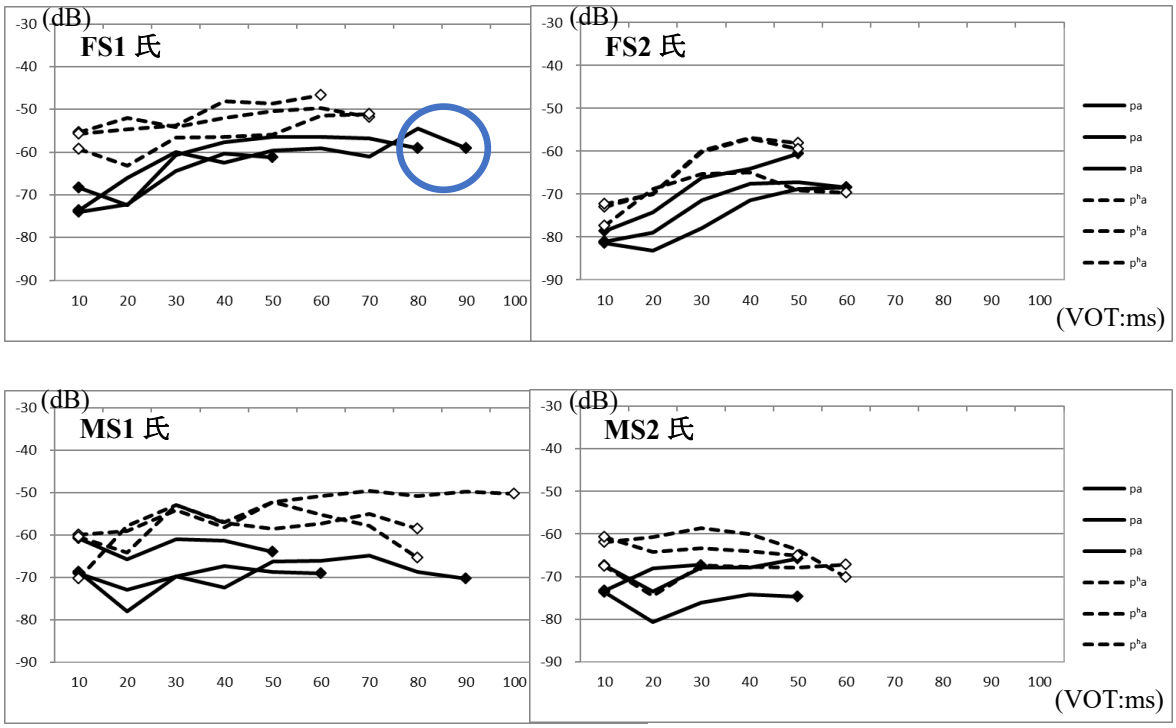


図 2:子音区間における高周波数帯域の噪音成分のパワー比較（平音/pa/:実線，激音/pʰa/:点線）

まず，FS1 氏（図 2：上段左グラフ）に注目すると，破裂時の 0～10ms から終了時まで，途中 50ms 時点では平音（実線）が-56.43dB，激音（点線）-55.85dB と差はわずかであるが，激音と平音のデータは重なることなく，常に「平音＜激音」を保っている． MS1 氏（下段左グラフ）では，10ms 時点では平音（実線）1 データが-60.82dB，激音（点線）の 1 データが-70.35dB と逆転しているが，20ms 以降ではデータが重なることなく「平音＜激音」を保っている．FS2 氏（上段右グラフ）と MS2 氏（下段右グラフ）でも，激音（点線）3 データのうち 2 データは平音（実線）のデータと重なることはなく，被験者 4 名すべてにおいて「平音＜激音」の傾向が観察される．これは，調音位置に関係なく，同様の結果である：

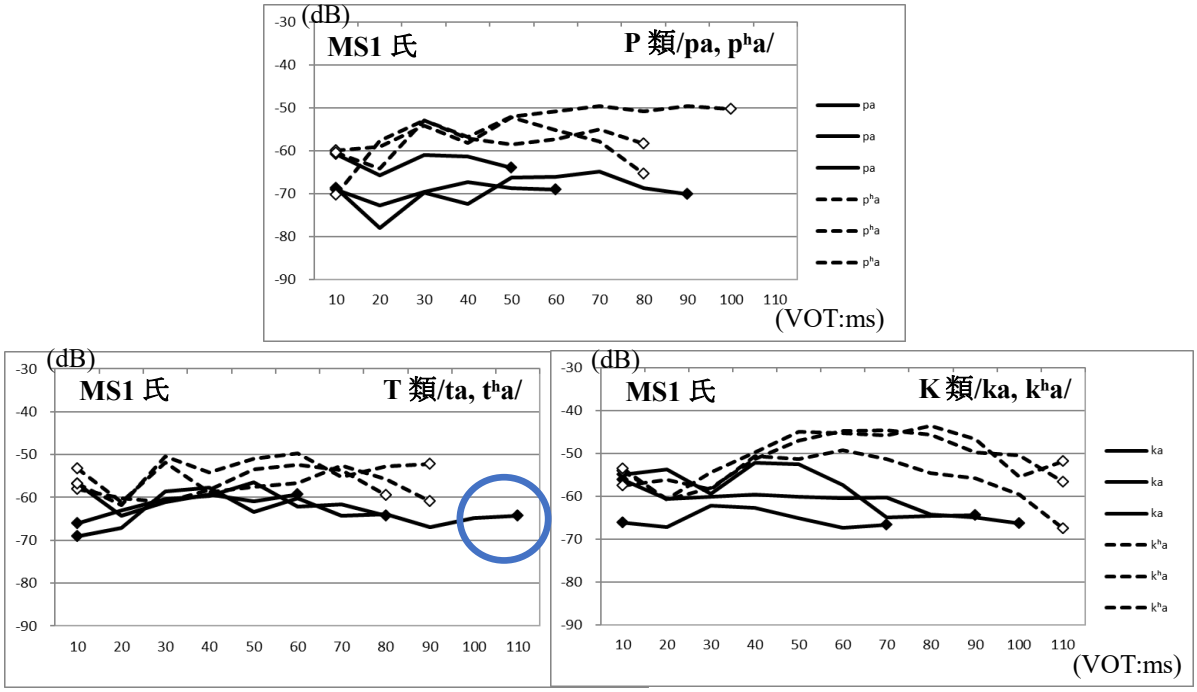


図 3:子音区間における高周波数帯域パワー比較(MS1 氏，平音:実線，激音:点線)

図 3 は MS1 氏による P 類/p, pʰ/（上段），T 類/t, tʰ/（下段左），K 類/k, kʰ/（下段右）の高周波数帯域の噪音成分のパワー比較グラフである．VOT が 10ms から 40ms ほどまでは，平音（実線）と激音（点線）が重なることがあるが，激音は VOT 後半に行くにつれて平音との差（平音＜激音）を保つかのように大きなパワーが持続していることがわかる．このような噪音成分のパワーの時間的な動きは，パワースペクトルでは捉えにくい．

3.2. 高周波数帯域の噪音成分のパワーと VOT との相関

興味深いのは，この高周波数帯域の噪音成分のパワーの大きさと VOT に相関がないことである．たとえば，図 2 の FS1 氏（上段左グラフ）と図 3 の MS 1 氏（下段左グラフ）の丸囲みに注目すると，横軸に示した VOT が激音（点線）に比べて，平音（実線）が長いにも関わらず，高周波数帯域の噪音成分のパワーは「平音＜激音」を保っている．

平音と激音では VOT 値が重複してきている（Silva 2006 ほか多数）ことは本発表のデータからも確認できた．しかし，VOT 値に差がない，または重複・逆転していても（VOT が激音≦平音），平音と激音では子音（VOT）区間の高周波数帯域の噪音成分のパワーが異なっている．パワーが異なるということは，調音時の口の構えが異なっており，よって産出される音色が異なっていることを意味している．VOT が平音より短い場合でも，激音であれば高周波数帯域の噪音成分のパワーが大きいことから，激音の説明に用いられることが多い「気息」「息を激しく，強く」といった音声的な特徴は，VOT よりも，高周波数帯域の噪音成分のパワーが平音よりも大きいこと，つまり音色の違い，子音の摩擦性が両者の対立に関与している可能性が高いことが指摘できる．よって，VOT 値に相関がなく，平音・

激音という対立で異なって観察される高周波数帯域の噪音成分のパワーという音響パラメータは，この両者の対立を支える，あるいは強化する（enhance）音響特徴であると言える．

4. おわりに

本発表では，ソウル方言における語頭の子音対立を支える音響特徴として，子音区間の高周波数帯域の噪音成分のパワーという音響パラメータを提案した．語頭が激音であればVOT の長さにかかわらず，平音よりもパワーが大きい．この結果を考慮すると，平音と激音の VOT 値の重複が進んでも，このパワーの現れ方の違いによって平音と激音の対立は保たれ続ける可能性がある．

4.1. 3 系列子音の対立を保つシステム

本発表で提案した，高周波数帯域の噪音成分のパワーを加え，3 つの子音系列（平音・激音・濃音）の対立システムを模式化にしたものが図 4 である．すなわち，「VOT」によって「濃音」と「それ以外」，「F0」によって「平音」と「それ以外」，今回提案した「高周波数帯の噪音成分のパワー」によって「激音」と「それ以外」というように，二項対立が組み合わさり，ソウル方言の 3 系列（平音，激音，濃音）の対立が維持される．この噪音成分のパワーの音響パラメータを設定することで，平音と激音の対立を強化し，3 系列の子音対立システムがよりよく説明できる．

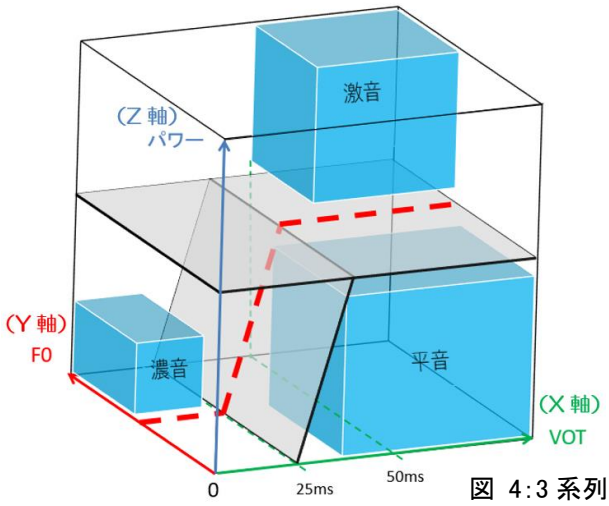


図 4:3 系列子音の対立システム

4.2. 今後の課題

この噪音成分の特徴はソウル方言のように，平音と激音では VOT 値が重複してきている方言のみに観察されるのだろうか．朝鮮語には慶尚道方言のように高低アクセントを持つ方言も存在する．今後，方言間で異なる高周波数帯域の噪音成分のパワーと VOT との相関を記述することは，方言類型論の音声学的基盤の構築に寄与できると考える．

【引用文献】

Kent, Ray D. and Charles Read (1992) *The Acoustic Analysis of Speech*, Singular Publishing Group.

Silva, David J. (2006) Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23, 287-308.

ケント，レイ・D / チャールズ・リード (1996; 1997)『音声の音響分析』荒井隆行・菅原勉 監訳，海文堂出版.

坂本 真一・蘆原 郁 (2016)『「音響学」を学ぶ前に読む本』コロナ社.

佐藤大和・益子幸江 (2013)「言語音声の聴知覚研究のためのツール構築」『語学研究所論集』第 18 号, 1-18, 東京外国語大学.

山崎亜希子 (2014)「ソウル方言における語中母音間破裂音の音響音声学的特徴－三項対立を支える音響特徴に関する考察－」『言語・地域文化研究』第 20 号, 121-133, 東京外国語大学大学院.