10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Articulatory asymmetry in consonantal sequences: a case from English, Fukui Japanese and Chaozhou Chinese

Kenji Yoshida[1], *Akira Utsugi*[2], *Wu Jia Hui*[3], *Tetsuo Nitta*[4], *Kiyoe Sakamoto*[1], *Yoko Ichimura*[5]

[1]Faculty of Humanities, Japan Women's University, Japan
[2]Nagoya University, Japan
[3]Graduate School of Humanity, Nagoya University, Japan
[4] Faculty of Letters, Kanazawa University, Japan
[5]National institute of Technology (KOSEN), Fukui College, Japan

kenjiyo.work@gmail.com, utsugi.akira@j.mbox.nagoya-u.ac.jp, wujiahuist@yahoo.co.jp, tnitta@staff.kanazawa-u.ac.jp, kiyoe.s@fc.jwu.ac.jp, ichiyo0307@gmail.com

Abstract

The present study examines the acoustic and articulatory characteristics of two types of consonantal sequences involving nasals and plosives: '(non-nasal) plosive + nasal' (involving 'nasal plosion') and 'nasal + plosive'. The nasal plosion data were recorded from two speakers of North American English (e.g., *hidden* [hɪdn]) and three speakers of the *Fukui* dialect of Japanese (e.g., [*mi tnta*] 'has already seen'): the coordination of oral and nasal closure/aperture observed with real-time MRI is similar between the two languages. The data for 'nasal + plosive' sequences were from 'denasalized' pronunciations of nasals from two speakers of Chaozhou Chinese (e.g., [mbik] 'honey'): the results for all speakers reveal similar events of articulation (simultaneous oral/velar closure and velar opening) for both nasal plosion and denasalization, but in the opposite order. However, the velar opening for nasal plosion is more abrupt and larger in volume, resulting in a long and loud nasal resonance. The opening is very small and brief in time for the denasalization cases, resulting in relatively weak nasality. The results indicate asymmetry in how oral and naso-pharyngeal articulations are organized in the two types of consonantal sequences, reflecting their different metrical status – syllabic vs. onset.

Index Terms: articulatory timing, nasal plosion, denasalization, real-time MRI

1. Introduction

Producing speech sounds involves motor control of relevant articulators. When producing sequences of speech segments, precise temporal coordination of multiple articulators is required [1]. Partly because of this challenge, certain types of segmental sequences are typologically rare [2], or more susceptible to historical changes or contextual variation. Producing nasal consonants involves coordination of oral and velic opening and closure. When closure of the former completely overlaps the opening of the latter, a nasal consonant is produced. If they are shifted in timing, on the other hand, a sequence of nasal and non-nasal consonant results [1, pp.320], [5, pp.228]. Instances of the second type can be found in, e.g., 'prenasalized stops' in Sinhala and Fula [3, pp.119-123] and in English words as *Thompson*. Sequences of the same sorts of consonants in the opposite order − 'prestopped nasals' or 'nasally released stops' − are known to be typologically less common [3, pp.128-129], especially word-finally [2, pp.166]. The present study investigates the characteristics of the sequences of nasal-stop and non-nasal stop in the two opposite orders, in terms in their acoustic characteristics and articulatory coordination.

2. The targets of the study

2.1. Nasal plosion

The production of nasal plosion has been described in terms of its articulation, which consists of two main phases: simultaneous oral and velic closures, and following release (lowering) of the velum [4, 5], which produces audible nasal murmur. Examples of this type of sound can be found in English (*hidden* [hɪdn], *kitten* [kɪtn], etc.), as well as for various genealogically unrelated languages [5, pp.362]. It has also been documented for the Fukui dialect of Japanese [6], [7], and the *Utai* singing in Japanese *Noh* theatrical plays [8], [9]. Except for our own studies [7-9], we have not found instrumental investigation into the production of this sound type. The description of the articulatory processes cited above was (presumably) obtained through introspection of the authors who are native speakers of English.

2.2. Prenasalized stops

According to the preceding literature [1, 3, 5], production of prenasalized stops involves the same articulatory phases as in the nasal plosion in the opposite order: velum lowering with oral closure, followed by simultaneous oral and velic closures, resulting in cessation of nasal murmur. This (alleged) articulatory symmetry is reflected in the common IPA transcriptions with the same elements in the opposite orders − [mb] and [bm]. Unlike nasal plosion, there have been numerous instrumental studies on prenasalized stops, particularly on the acoustic and aerodynamic characteristics of prenasalized stops arising through "denasalization" of simple full-nasal consonants [10-15].

2.3. The goal of the present study

This paper reports some of the results from the project of the authors investigating production of consonantal sequences with a nasal consonant being one of their elements. One of the goals is to examine the articulatory processes in producing sequences of segments. Nasal plosion and prenasalized stops were chosen as the focus of the study because they require coordination of two articulators that are relatively independent

10.21437/SpeechProsody.2020-5

from each other (oral and velic). The results from acoustic and articulatory data analysis will be reported on two languages having nasal plosion – English and Fukui Japanese – and another having prenasalized stops – Chaozhou Chinese. The findings reveal a stark difference in the span and the magnitude of the velic opening, demonstrating articulatory asymmetry between the two consonantal sequences.

3. Acoustic properties

3.1. Nasal plosives

First, the results from the acoustic analyses are drawn from the previous study of the authors, where the data from two North-American speakers of English and three speakers of Fukui Japanese were analyzed [7, 8]. Figure 1 shows an example by a Canadian male speaker of English, producing the word *hidden* [hɪdn] in a carrier sentence. The segment representing the oral closure (annotated as [d]) is directly followed by a nasal segment ([n]). One can hear nasal plosion between the two segments, where an abrupt rise in amplitude is seen and a distinct resonance in the higher frequency region begins.

Figure 2 shows an example of nasal plosion by a male speaker of Fukui Japanese in the resultative form (the structure where nasal plosion is typically observed in this dialect) of a verb *miru* 'look'. As in English, the segment of oral closure ([t]) is followed by the segment with nasal resonance ([n]). Between the two segments, a brief period of aspiration ([h]) is observed, suggesting velum lowering (thus, airflow through the nostrils) but that the vocal cord vibration and the resonance in the naso-pharyngeal port have not started. This kind of time lag is less frequent but observed for English speakers as well [8, pp.51].

Overall, the observations demonstrate that pronunciations involving nasal plosion in English and Fukui Japanese are quite similar in their acoustic characteristics. Their similarity will be reconfirmed in their articulation in Section 4.

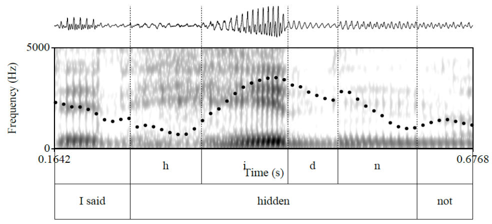

Figure 1: *An example of nasal plosion by a male Canadian speaker of English ('hidden'): Sound waveform (top), Spectrogram with fo contour (middle) and annotations (bottom).*

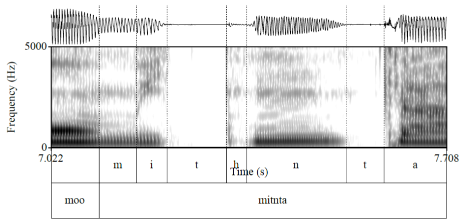

Figure 2: *An example of nasal plosion by a male speaker of Fukui Japanese.* ([mo: mi tnta] 'I have already seen (it).')

3.2. Prenasalized stops

Examples for pre-nasalized stops are drawn from the speech data from two speakers of Chaozhou Chinese: Speaker 1, from *Shantou*-city; Speaker 2, from *Jieyang*-city (the cities are next to each other). They produced 18 and 17 monosyllabic words, respectively, in a frame sentence '這 個 字 讀 __ 吧 (This letter reads ___, doesn't it?)'. The experimental words had bilabial ([m] or [mb]), dental ([n]), or velar ([ŋg]) consonant as the onset. Figure 3 shows a sample of word 文 'letter' by Speaker 2. The acoustic events unfold in the opposite order from what we have seen for English and Fukui Japanese: the segment with nasal murmur followed by the segment with (mostly) voice bar alone, suggesting cessation of nasality.

While a considerable variation in pronunciation has been reported among Min Nan Chinese [10, 12], the two speakers of the present study mostly agreed in their pronunciation of the experimental words. Speaker 1 pronounced all the 10 words with /m/ in the word-initial position with a pre-stopped nasal (Figure 4, left). Speaker 2, however, pronounced 2 of these words (焖 and 密) with a full nasal consonant, providing two minimal pairs with a full nasal vs. pre-nasalized stop (焖 vs. 文, and 密 vs. 蜜).

To sum up so far, we have seen that nasal plosion in English and Fukui Japanese, and pre-nasalized stops in Chaozhou Chinese show a fairly good symmetry in their acoustic characteristics. We will examine if this applies to their articulatory organization as well.

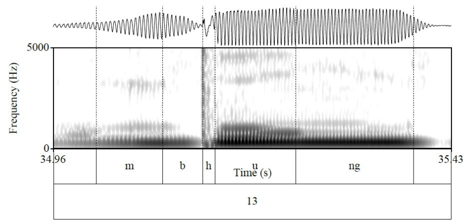

Figure 3: *An example of pre-nasalized stop by a female speaker of Chaozhou Chinese* (Speaker2*:* [mb^{h}uŋ] 'letter').

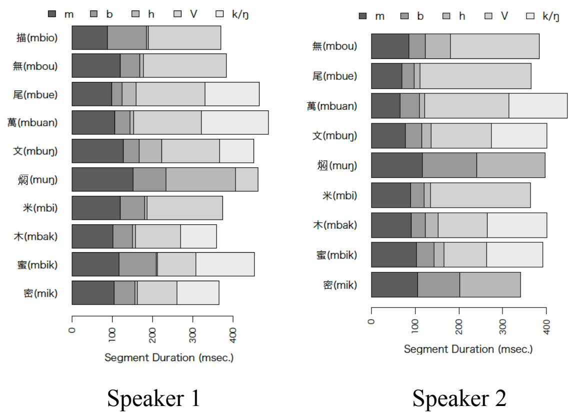

Figure 4: *Mean duration (in msec.) of acoustic segments of the experimental words with /m/ as the word-initial consonant.*

4. Articulatory Processes

4.1. Methods

To understand the articulatory coordination of the consonantal sequences in question, the time-varying patterns of movements of the articulators must be observed. Real-time MRI (Magnetic Resonance Imaging) movies were recorded for the speech organs of the speakers at the ATR (Advanced Telecommunication Research Institute International), with the operation and technical support of BAIC (Brain Activity Imaging Center). The speakers lay inside the MRI machine and read aloud the experimental words embedded in a short frame sentence (e.g., English '*I said* X, *not* Y.'). In each of the MRI sessions (30 seconds), the speaker produced the same utterance repeatedly (7 to 14 times). A sequence of images at the mid-sagittal section of the speaker's head was thus obtained (Figure5; 256 × 256 mm; resolution; 1 × 1 × 10 mm). Each speaker completed 8 to 18 sessions in no more than 50 minutes. This procedure had been approved by the ATR Examination Board for the Safety of fMRI, MEG Studies.

The MRI images were adjusted for brightness and contrast and the outlines of relevant articulators were manually drawn and recorded using ImageJ [16]. For English and Fukui Japanese, the outlines of the tongue, the palate (including velum) and the rear pharyngeal wall were drawn to estimate the magnitudes of the oral and velic aperture. For Chaozhou Chinese, the outlines of the upper and lower lips and the palate were drawn. The target frames were identified referring to the audio track recorded simultaneously with the MRI recordings (the scanner noise was reduced on Praat [17]). The real-time MRI system at the ATR provides the time resolution of 10 frames/second, which is coarser than, e.g., the one developed in the University of Göttingen [18] (18 frames/sec.), because of the trade-off for the finer spatial resolution. This is not sufficient to capture very rapid changes in configuration of the articulators. It is difficult to precisely determine when two articulators make a contact and it is likely that corresponding frames in the repeated renditions do not exactly fall on the same timing. To overcome these methodological limitations, the mean values were computed for oral/velic opening across the repeated renditions for each of the corresponding frames. By doing this, the shifts in timing among the frames would be cancelled out to a certain extent and a rough approximation of the time-varying pattern of articulatory configuration can be obtained. Sample of the MRI tracings will be uploaded at Nagoya Repository (https://nagoya.repo.nii.ac.jp/).

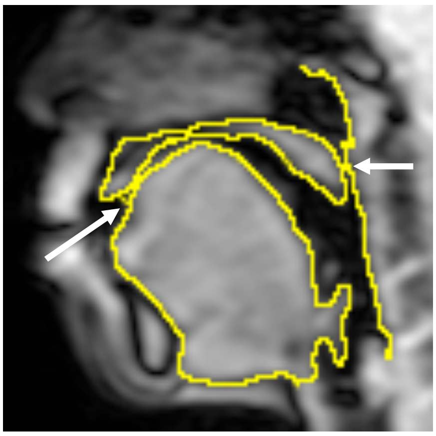

Figure 5: *An example of MRI tracing: a female speaker of Fukui Japanese* ([mi tnta]', 1st frame). *The tongue tip aperture* (TT) *is* 3.89 *(marked by a 45 degree line) and the velic aperture* (VEL) *is* 2.00 *(the horizontal line), both indicated by the arrows.*

4.2. Articulation of nasal plosion

For English nasal plosion, the MRI movies on 4 words (*hidden*, *bitten*, *eaten*, *forbidden*) located in two prosodic positions (Focus/Post-focus) were recorded for four speakers. The distance between hard palate and tongue blade (the minimal distance at 45 degrees for consistency) was measured for oral closure/opening, and the minimal horizontal distance between velum and the rear pharyngeal wall was measured for the velic closure/opening. The result for *hidden* in Focus position for a female North-American speaker is shown in Figure 6. This shows the mean values for each of the 6 frames, which roughly corresponds to the underlined part of the utterance, '*I said* <u>*hidden, no*</u>t *hitting*.' The aperture of 0.0 indicates the contact (i.e., closure) at the alveolar region (TT) or at the velum (VEL). The larger the aperture (higher in the vertical axis), the greater the distance between the two articulators. The time course of TT and VEL conforms to the articulatory configuration described in the literature [4, 5]. Simultaneous oral and velic closures is achieved around Frame 3, which corresponds to the non-nasal stop ([d]). Toward Frame 4 (and 5), the velum lowers while the alveolar closure is maintained, which corresponds to the resonance in naso-pharyngeal port ([n]).

For Fukui Japanese, the MRI movies for 10 verb phrases (*mi tnta* (have seen), *ne tnta* (slept), *tabe tnta* (eaten), etc.) were recorded for three speakers. Figure 7 shows a case from a male speaker, where the mean TT and VEL values over 8 frames, corresponding to the underlined part of the utterance '*Moo* <u>*ne tnta*</u>.' are plotted. The time course of articulatory configuration closely resembles that of English. The simultaneous oral and velic closure is at around Frame 3, followed by nasal release toward Frames 4 and 5.

Essentially the same pattern is observed in Figure 8, which shows the case for '*Moo* <u>*mi tnta*</u>' by a female speaker of Fukui Japanese (the speaker for Figure 5), with the oral/velic closure at around Frame 3 and nasal release at around Frames 4 and 5. A similar pattern is also observed for Japanese *Noh* singing, albeit it unfolds in a much longer time span [9].

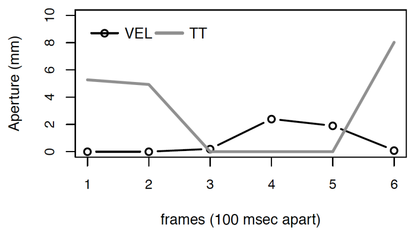

Figure 6: *The mean aperture for tongue tip* (TT) *and velum* (VEL) *for a female American English speaker* (*hidden*; N=10).

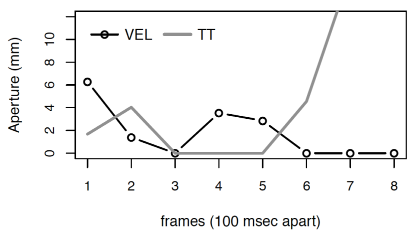

Figure 7: *The mean aperture for* TT *and* VEL *for a male speaker of Fukui Japanese* ([ne tnta]; N=11).

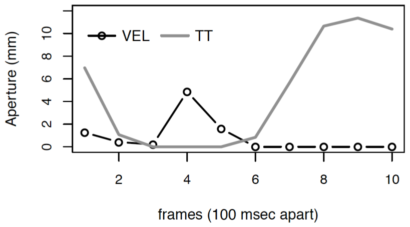

Figure 8: *The mean aperture for* TT *and* VEL *for a female speaker of Fukui Japanese* ([mi tnta]; N=9).

Overall, the results indicate that nasal plosion is produced with the timing coordination of oral and velic articulators largely as described in the literature. Moreover, the pattern observed seems to be general across the two languages.

4.3. Articulation of prenasalized stops

As we have seen in section 3.2, prenasalized stops in Chaozhou Chinese have a minimal (full-nasal) counterpart for Speaker 2 (e.g., [mik] in contrast to [mbik]). This is a consequence of a historical change whereby nasal consonants in a preceding stage of the language split into prenasalized stops, which is particularly the case for Chaozhou Chinese [10, pp.90-91]. Figure 9 shows a case for full-nasal 焖 [muŋ] 'simmer': the mean labial aperture (LAB) and velic aperture (VEL) over 9 frames, which corresponds to the underlined part of the utterance [tsi kai zi t<u>ak muŋ pa</u>] (see section 3.2 for the frame sentence written in Chinese characters). As expected, the timing configuration of the two articulators is totally different from what we have seen for English and Fukui Japanese. The velic opening is maintained throughout most of the test word, with anticipatory nasalization around Frame 1 and nasalization of the tauto-syllabic vowel, maintained up to the syllable-final [ŋ] at around Frame 6.

Figure 10 shows the result from the minimal (denasalized) counterpart 文 [mbuŋ] 'letter'. The pattern is again different from the case of full-nasal onset. The velum is quickly closed from Frame 2 toward Frame 3, indicating cessation of nasality, with the labial closure maintained. Although this exhibits the sequence of the same articulatory configuration in the opposite order from what we have seen in nasal plosion, there is a significant difference between the two. For nasal plosion, the opening of the velum is large in magnitude (Figure 6-8; max. 2.4, 3.5, 4.9 mm, respectively) and long in duration. For prenasalized stops, the velum opening is very small in magnitude (Figure 10; max. 0.1 mm) and brief in time. A similar pattern is observed for the same test word from Speaker 1. The comparison of nasal plosion in English and Fukui Japanese with prenasalized stops in Chaozhou Chinese indicates that the two consonantal sequences are not the mirror images of each other in their articulatory organization.

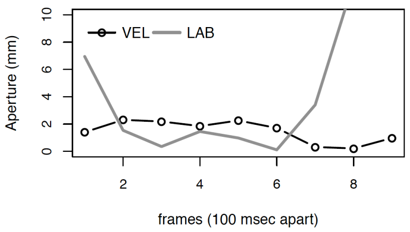

Figure 9: *The mean aperture for* LAB *and* VEL *for Speaker 2 of Chaozhou Chinese* ([muŋ] 'simmer'; N=9).

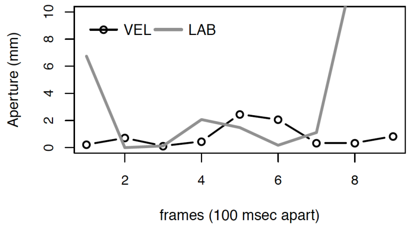

Figure 10: *The mean aperture for* LAB *and* VEL *for Speaker 2 of Chaozhou Chinese* ([mbuŋ] 'letter'; N=9).

5. Discussion and conclusions

Comparison of two types of consonantal sequences with the same elements in different orders, 'stop+nasal (nasal plosion)' and 'nasal+stop (prenasalized stops)' in three languages reveals a substantial difference in the organization of relevant articulatory processes. The former involves the velic opening with a large magnitude and long duration, whereas the latter has a very small and brief period of velic opening. Shifts in articulators in supine position have been reported [19, 20], which could influence the results. However, given that the shifts are particularly smaller for the targets of consonants [19, pp.539], we believe that the observed difference between the two consonantal sequences is reliable. This situation is very unlike the "mirror images" as suggested, e.g, by [5, pp.228; Fig. 8.5, (b) and (c)]. Articulatory configuration of nasals is different depending on the positions in a syllable [21]. The present finding provides further demonstration of the way speech production can be conditioned by the metrical status of the speech segments − syllabic vs. onset.

The differences observed in the present study may also be rooted in the difference in their historical sources. Nasal plosion typically arises through vowel syncope [22, pp.107], resulting from shortening of stressless vowels [23, pp.34]. A similar process must have been at work for the nasal plosion in Fukui Japanese, as observed by Nitta [6]. Although it is masked by the oral closure, the original articulatory/acoustic configuration involved a 'release', which can be manifested as an abrupt and large naso-pharyngeal resonance with the coordination of articulations as observed here. On the other hand, the source of prenasalized stops is weakening of nasality at the syllable-onset position [11, 14]. The smaller and briefer nasalization in the onset position can be understood in relation to the finding that it is a 'strong' position that is more susceptible to prosodically conditioned acoustic [13] and articulatory enhancement [21, pp.50], [24]. Although the human auditory system can detect a subtle nasality, it may be the case that nasality can be suppressed to a great extent in certain metrical positions. Perceptual investigation of speakers of these languages is important for for further elucidation of the relevance of the present findings to the nature of temporal organization of speech.

6. Acknowledgements

We are grateful to the speakers for their cooperation to the data collection. We also thank to the technical support and advice on the MRI movies from the members of BAIC (https://baic.jp/) and from Tatsuya Kitamura. The study reported here has been supported with the Research Grant by Japan Society for the Promotion of Science (17K02692).

7. References

[1] C. Browman and L. Goldstein, “Gestural structure, distinctiveness, phonological processes, and historical change,” in *Modularity and the Motor Theory of Speech Perception*. New York: Academic Press, 1991, pp. 313-338.

[2] E. Moravscik, *Introducing Language Typology*. Cambridge: Cambridge University, Press., 2013.

[3] P. Ladefoged and I. Maddieson, *The Sounds of the World’s Languages*, Cambridge, MA: Blackwell Publishers, 1996.

[4] P. Ladefoged and K. Johnson, *A Course in Phonetics*, *6th edition*. Boston: Wadsworth, 2010.

[5] J. Laver, *Principles of Phonetics*. Cambridge: Cambridge University Press, 1994.

[6] T. Nitta, “Highlights of Japanese dialects: Fukui Prefecture,” *Nihon Gogaku* vol. 34, no. 4, 2015, pp. 78-79.

[7] K. Yoshida, T. Nitta, Y. Ichimura, and A. Utsugi, “Nasal plosive in Fukui Japanese: the pattern of its segmental duration,” *Proceedings of the 32nd Annual Meeting of Phonetic Society of Japan*, 2018, pp. 144-149.

[8] K. Yoshida and K. Sakamoto, “Preliminary analysis on the production of nasal plosion: a case from English and Japanese *Noh* singing,” *Ronshu XII*, 2017, pp. 47-63.

[9] K. Yoshida, and K. Sakamoto, “Nasal plosion in Japanese *Noh* singing: its acoustic and articulatory characteristics,” *Proceedings of the 33rd Annual Meeting of Phonetic Society of Japan*, 2019, pp. 198-203.

[10] M. Chan, “Post-stopped nasals in Chinese: An areal study,” *UCLA Working Papers in Phonetics,* vol. 68, 1987, pp. 73-119.

[11] M. Chan and H. Ren, “Post-stopped nasals: an acoustic investigation”. *UCLA Working Papers in Phonetics,* vol. 68, 1987, pp. 120-129.

[12] F. Hu. “Post-oralized nasal consonants in Chinese dialects – aerodynamic and acoustic data,” in *International Congress* of *Phonetic Sciences XVI*, 2007, pp. 1405-1408.

[13] K. Yoshida, “Phonetic implementation of Korean “denasalization” and its variation related to prosody,” *Indiana University Working Papers in Linguistics Online*, vol.8, 2008, pp. 1-20.

[14] P. S. Ding, “The partially denasalized bilabial plosive in Southern Min: comparison to [mb] in Amdo Tibetan, ” *International Congress of Phonetic Sciences XVII*, 2011, pp. 584-587.

[15] Y. S. Kim, “*An Acoustic*, *Aerodynamic and Perceptual Investigation of Word-initial Denasalization* in *Korean*,” Doctoral dissertation, University College of London, 2011.

[16] C. A. Schneider, W. S. Rasband and K. W. Eliceili, “NIH image to ImageJ: 25 years of image analysis.” *Nature Methods*, vol. 9, no. 7, 2012, pp. 671-675.

[17] P. Boersma and D. Weenink, “Praat: doing phonetics by computer. Version 6.0.43”, 2018, computer program retrieved from http://www.praat.org/.

[18] M. Ueckera, S. Zhanga, D. Voita, A. Karausa,K-D. Merboldta and J. Frahm, “Real-time MRI at a resolution of 20 ms.” *NMR in Biomedicine*, vol. 23, 2010, pp.986-994.

[19] M. Stone, G. Stock, K. Bunin, K. Kumar, M. Epstein, C. Kambhamettu, M. Li, V. Pathasarathy and J. Prince. “Comparison of speech production in upright and supine position.” *Journal of the Acoustical Society of America*, vol. 122, no. 1, 2007, pp.532-541.

[20] T. Kitamura, H. Takemoto, K. Honda, Y. Shimada, I. Fujimoto, Y. Syakudo, S. Masaki, K. Kuroda, N. Oku-uchi and M. Senda, “Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner.” *Acoustical Science and Technology*, vol. 26, no. 5, 2005, pp.465-468.

[21] R. Krakaw, “Physiological organization of syllable: a review,” *Journal of Phonetics*, vol. 27, 1999, pp. 23-54.

[22] M. Gordon, *Phonological Typology*. Oxford: Oxford University Press, 2016.

[23] J. Bybee, *Language Change*. Cambridge: Cambridge University Press, 2015.

[24] T. H. Cho and P. Keating “Articulatory and acoustic studies on domain-initial strengthening in Korean,”, *Journal of phonetics*, vol. 29, 2000, pp. 155-190.