Perception of Korean nasal onset /m/ by Japanese listeners: A preliminary study

Jihyeon Yun^{1,2,*} and Takayuki Arai¹

¹Graduate School of Science and Technology, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan

²Graduate School of Linguistics, Chungnam Nat'l University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134 Republic of Korea

(Received 1 March 2018, Accepted for publication 18 June 2018)

Keywords: Nasal consonant, Denasalization, Nasal weakening, Duration, Intensity
PACS number: 43.71.Es [doi:10.1250/ast.39.424]

1. Introduction

Seoul Korean speakers tend to pronounce word-initial nasal consonants /m/ and /n/ with weak nasal airflow and weak nasal energy [1,2]. Along with the devoicing of oral stops, the nasals of northern dialect are reportedly characterized by occasional release burst, less voicing and nasal murmur, and shorter duration [3] in the domain initial position, with inter- and intra-speaker variation. The degree and frequency of this tendency vary depending on the speakers' dialect and age group [1,4].

While Korean word-initial nasals can be perceived as oral stops by non-native listeners, the native speakers have no difficulties recognizing the "denasalized" /m/ and /n/ as genuine nasals, regardless of whether the speakers themselves denasalize their nasals.

This intriguing dichotomy can be understood within the phoneme inventory and phonology of Korean: while the only possible competitor as an oral voiced stop is a lax stop, this consonant is devoiced and aspirated in the domain initial position. Therefore, a [b]-like nasal and a [p]-like lax consonant do not compete with each other, that is, denasalized or devoiced onset nasals are not confused with other oral stops [5].

The present research investigated the perception of Korean onset /m/ by Japanese listeners using naturally spoken monosyllables as well as manipulated stimuli with a shortened onset consonant interval. The duration and the intensity of the nasal were analyzed together with a perception test.

2. Materials

2.1. Recording

The original speech sounds were recorded in a soundproof room located at Sophia University, Japan. A female Chungnam Korean speaker (in her 30s) repeatedly pronounced isolated /ma/, /me/, /mi/, /mo/, and /mu/ in isolated form. The utterance was recorded with a dynamic microphone (TOA DM-1300) connected to a digital PCM recorder (Marantz PMD661MKII) with a sampling rate of 48 kHz.

2.2. Sound evaluation

The second author, who is a native speaker of Tokyo Japanese, evaluated the recording and judged that most of the

/mi/, /mo/, and /mu/ samples sounded like [bi], [bo], and [bu], that is, as if those had oral voiced plosive onsets (Table 1). Since presenting all those sounds in the experiment would not only be redundant but might also induce response bias, /mi/, /mo/, and /mu/ samples were eliminated from the stimulus set.

Ten tokens from each of the /ma/ and /me/ recordings were selected as the experimental stimuli based on the evaluation. Since the second author judged most of the stimuli as denasalized, nasal-like tokens were preferentially included in the stimulus set, in order to prevent the set from consisting only of oral-like samples.

2.3. Manipulation of onset duration

To investigate whether the duration of nasal affects the perception for Japanese listeners, we created stimuli with different onset durations in Praat version 6.0.17 [6]. For manipulation, nasal murmur in the onset consonant of each syllable was selected (murmurC).

The start of murmurC was set at the beginning of pulsing, on the zero crossing point with a negative slope in the waveform. The endpoint of murmurC was decided based on the visual inspection of the start of the transition for the first two formants (F1 and F2) in the Sound Editor window of Praat [6]. This endpoint was set at or before the time point where the formant tracks (mostly F1) in the Sound Editor of Praat [6] began to rise, so that information on the consonantality and the place of articulation were auditorily retained after removal of murmurC interval.

The murmurC interval of the original recordings (step0) was shortened in two steps:

- (i) step1 to half of the original duration
- (ii) step2 to 0 seconds, which resulted in the removal of nasal murmur interval

This manipulation was performed with the duration manipulation function in Praat version 6.0.17 [6]. Table 2 shows a summary of the statistics for duration of murmurC (unit: ms) for the original sounds (step0). The individual duration of murmurC for each stimulus are shown in Table 3 below with the result of the experiment.

3. Perception experiment

3.1. Stimuli

The stimulus set consisted of 60 isolated syllables originally derived from Korean /mV/. The stimuli differed

*e-mail: yunjihyeon@gmail.com

Table 1 Number of sound samples (*n*), mean, standard deviation (*sd*), and range of duration for voiced intervals of /m/ (in ms), and number of samples judged as *nasal* by the second author.

	<i>n</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>n(nasal)</i>
/ma/	16	51.2	20.2	81.1	5
/me/	16	44.6	24.4	87.1	8
/mi/	16	59.6	17.4	59.8	0
/mo/	16	53.8	28.5	125.6	0
/mu/	14	54.9	22.8	72.7	0

Table 2 Duration of murmurC (unit: ms) for the original sounds (step0).

	<i>min</i>	<i>max</i>	<i>median</i>	<i>sd</i>
/ma/	17	103	50	24.2
/me/	16	92	52	23.6

Table 3 Duration of murmurC interval (in ms) and percentage of response as /mV/ (%nasal) are shown for each step of duration manipulation (step). Values are rounded off to the nearest whole number.

/Ca/ source		step			/Ce/ source		step		
		0	1	2			0	1	2
ma01	murmurC	76	38	0	me01	murmurC	74	37	0
	%nasal	38	44	31		%nasal	69	56	31
ma02	murmurC	103	51	0	me02	murmurC	92	46	0
	%nasal	0	0	0		%nasal	44	19	25
ma03	murmurC	64	32	0	me03	murmurC	49	25	0
	%nasal	0	0	0		%nasal	75	81	69
ma04	murmurC	45	23	0	me04	murmurC	38	19	0
	%nasal	75	94	69		%nasal	6	12	12
ma05	murmurC	34	17	0	me05	murmurC	60	30	0
	%nasal	50	44	44		%nasal	100	88	69
ma06	murmurC	34	17	0	me06	murmurC	50	25	0
	%nasal	0	0	0		%nasal	100	69	38
ma07	murmurC	50	25	0	me07	murmurC	54	27	0
	%nasal	50	69	31		%nasal	88	88	75
ma08	murmurC	50	25	0	me08	murmurC	61	30	0
	%nasal	0	6	6		%nasal	94	94	81
ma09	murmurC	58	29	0	me09	murmurC	16	8	0
	%nasal	19	25	6		%nasal	81	75	56
ma10	murmurC	17	8	0	me10	murmurC	16	8	0
	%nasal	62	50	25		%nasal	94	88	4
mean pooled	murmurC	53	27	0	mean pooled	murmurC	51	26	0
	%nasal	29	33	21		%nasal	75	67	55

in the original sound (source) and the ratio of shortening for onset interval by duration manipulation (step). There were 30 stimuli of /Ca/ and another 30 of /Ce/.

3.2. Participants

Eight normal-hearing Japanese listeners (7 females and 1 male) participated in the perception experiment. Their ages varied between 20s and 40s (years, *min* = 22, *max* = 43, *mean* = 28.5, *sd* = 6.9).

3.3. Procedures

The perception experiment was conducted in a sound-proof room at Sophia University, in December 2016 and January 2017. The stimuli were presented through headphones (Audio-Technica ATH-A900) via the audio interface (Edirol UA-25EX) connected to a laptop computer (Lenovo ThinkPad T410, Windows 8.1). The experimental platform was Praat ExperimentMFC [6].

The instructions were provided in Japanese. Under the forced choice classification design, a single stimulus was presented in each trial in random order but not the same stimuli consecutively. The stimulus of a subsequent trial was played automatically 2 seconds after the previous response.

The choices appeared on the computer screen in Japanese Katakana, “ma” and “ba” for the trials of /Ca/ stimuli, and “me” and “be” for /Ce/ stimuli. The order of buttons was counter-balanced for each participant: the same stimulus appeared in two trials in nonconsecutive order, once with the /mV/ button on the left side of the screen, then the /bV/ button on the left in another trial.

The session for each participant consisted of a practice block and a main experimental block. Both /Ca/ and /Ce/ type stimuli were presented in mixed order within a block. The practice block comprised eight trials with /ma/, /ba/, /me/, and /be/ sounds spoken by a female speaker of Tokyo Japanese (in her 20s). The practice stimuli were recorded in the same environment except for the microphone (SONY ECM-23F5). The percentage of correct responses for the practice block was 97% for the pooled data.

In the main experiment, a “pause” screen appeared after every 20 trials so that the participants could take a break. It took approximately 15 minutes for a participant to complete the whole session.

3.4. Results

The data were obtained from a total of 960 trials (2 vowel environments × 10 original sounds × 3 steps of manipulated duration × 2 button orders × 8 participants). The percentage of response as /mV/ (%nasal) is shown in Table 3 with duration of murmurC interval for each stimulus.

The result did not show clear relationship between duration of the manipulated onset interval (murmurC) and perceptual judgment. Even when the duration and intensity of murmurC for step2 were both zero, some step2 stimuli were still perceived as nasal by the listeners.

4. Discussion

The stimuli of the experiment can be divided into three types according to the perceptual judgment:

- the original sound (step0) is perceived as oral by the listeners
- the original is heard as nasal, whereas the shortening of its consonant interval (step1 and step2) turns it to an oral in perception
- the original sound is perceived more or less nasal, and

Fig. 1 Example of a stimulus with the boundary for duration manipulation (murmurC) and with the modified CV boundary for intensity analysis (durC).

still nasal even after removal of the murmur interval (step2)

The percentage of responses as nasal was generally reduced when the nasal consonant interval was shortened or removed (hence, type (b)). Yet, some of the step2 stimuli with almost no consonant interval were judged to be more nasal than the stimuli with longer consonant duration.

When we selected murmurC interval for stimuli manipulation, one or two periodic waves, in which the formant tracks show transition before the location of burst-like strong energy between the consonant and its succeeding vowel, remained outside of the murmurC interval, and therefore they were kept unchanged in all stimuli. Thus we suspected that these periodic waves may contain critical information.

To test this idea, we remeasured the duration and the intensity of onset interval for each stimulus in Praat [6] including these periodic waves. The CV boundaries were set based on a noticeable change in intensity on the spectrogram. As a result, these boundaries were generally placed later in time than the previous endpoints of murmurC interval (Fig. 1).

To measure the intensity, each target segment (both C and V) was extracted, and silent intervals of 500 ms were concatenated before and after the target. Sound intensity maxima (maxIntC) and mean (meanIntC) were measured in dB for onset intervals (C) and the vowels (V).

Two parameters, maxIntDiff and meanIntDiff were calculated by subtracting the maximum intensity of vowel (maxIntV) from the intensity of its tautosyllabic onset as shown in (1) and (2) below.

$$\text{maxIntDiff (dB)} = \text{maxIntC} - \text{maxIntV} \quad (1)$$

$$\text{meanIntDiff (dB)} = \text{meanIntC} - \text{maxIntV} \quad (2)$$

Note that the reference value for the vowel interval is the maximum intensity (maxIntV) for both cases. The descriptive statistics are shown in Table 4.

Table 4 Summary of statistics for maxIntDiff and meanIntDiff (unit: dB).

	<i>min</i>	<i>max</i>	<i>mean</i>	<i>sd</i>
maxIntDiff	-15.59	-3.91	-7.92	2.70
meanIntDiff	-16.35	-6.37	-9.46	2.34

Fig. 2 Percentage of responses as nasal (%nasal) by the difference between intensity of consonant and tautosyllabic vowel (meanIntDiff): grey lines connect stimuli of same original sound source with different manipulation step.

The measured duration (durC), maximum intensity (maxIntDiff), mean intensity (meanIntDiff) of the onset consonants, and percentage of responses as nasal (%nasal) were not normally distributed. The correlation between two variables was inferred by Spearman's ρ (ρ) in R version 3.4.3.

Spearman's rank-order correlation tests were run to determine the relation between %nasal and each of the three parameters, namely, durC, maxIntDiff, and meanIntDiff for the 60 stimuli (2 vowels $\times$ 10 tokens $\times$ 3 steps of duration manipulation).

No correlation was shown between durC and %nasal [$\rho(58) = -0.005, p = 0.969$]. In contrast, a positive correlation was shown between maxIntDiff and %nasal [$\rho(58) = 0.643, p < 0.0001$], and an even stronger correlation was seen between meanIntDiff and %nasal [$\rho(58) = 0.718, p < 0.0001$]. %nasal by meanIntDiff is shown for each stimulus in Fig. 2.

It seems that the beginning of the vowel interval contains more information on nasality than the longer consonant interval with weak intensity and less nasal coupling does. As shown in Fig. 2, the stimuli with the greater intensity difference between consonant and vowel interval were more likely to be perceived as oral by Japanese listeners. This is contrastive to the fact that the relations between duration and perception did not show any significance.

On the other hand, reports regarding the influence of the tautosyllabic vowels on Korean nasal weakening have been

incongruous, e.g., more denasalization for the nasal onsets with high vowels [7], or no consistent effect of the vowel height [2]. In the present study, the authors judged the recorded monosyllables with high vowels to be more denasalized in general (Table 1), except that /me/ was more nasal than /ma/. Whether this was resulted from a certain feature of the mid front vowel /e/ or merely due to individual traits of the speaker is not clear from the current data.

5. Conclusion

The /mV/ syllables spoken in isolation by a Korean speaker had great variance in their duration of nasal interval. The percentage of nasal onset perception generally decreased when the onset duration was artificially shortened or removed, yet the duration of murmur itself was not significantly related to the perception.

The present study revealed that perception of Korean nasal consonant by Japanese listeners relies more on the intensity of the onset and presumably the quality of the information contained within that interval rather than the absolute duration.

Acknowledgements

The contents of the present report were presented at the Spring Meeting of ASJ in 2017.

References

- [1] K. Yoshida, "Phonetic implementation of Korean "denasalization" and its variation related to prosody," *IULC Working Papers Online*, **8**, 1–23 (2008).
- [2] Y. S. Kim, *An Acoustic, Aerodynamic and Perceptual Investigation of Word-initial Denasalization in Korean*, PhD diss., University College London (2011).
- [3] M. J. Ahn, "Acoustic duration of Korean nasals," *Stud. Phon. Phonol. Morphol.*, **19**, 411–431 (2013).
- [4] K. Yoo, "Domain-initial denasalization in Busan Korean: A cross-generational case study," *Proc. ICPHSXVIII* (2015).
- [5] T. Y. Kim, *Devoicing and Denasalization in Korean: Reconsideration of Phonological Structure in Korean*, MA thesis, Waseda University (2015).
- [6] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glott Int.*, **5**(9/10), 341–345 (2001).
- [7] M. Chen and H. Clumeck, "Denasalization in Korean: A search for universals," in *Nasalfest: Papers from a Symposium on Nasals and Nasalization*, C. A. Ferguson, L. M. Hyman and J. J. Ohala, Eds. (Stanford University Linguistics Dept., CA, 1975), pp. 125–131.