Korean and Japanese listeners' perception of /ma/ produced by a mechanical vocal-tract model

Jihyeon Yun[1,2,*], Janice Wing Sze Wong[3], Jeff Moore[4], Junya Iwakami[5], C. T. Justine Hui[1,6], Jiayin Gao[5,7] and Takayuki Arai[5]

[1]*Graduate School of Science and Technology, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan*
[2]*Graduate School of Linguistics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134 Republic of Korea*
[3]*Department of English Language and Literature, Hong Kong Baptist University, 224 Waterloo Road, Kowloon Tong, Hong Kong SAR*
[4]*Department of Information Networking for Innovation and Design, Toyo University, 1 Chome-7-11 Akabanedai, Kita-ku, Tokyo, 115–0053 Japan*
[5]*Department of Information and Communication Sciences, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan*
[6]*Department of Mechanical Engineering, University of Auckland, 20 Symonds Street, Auckland 1142 New Zealand*
[7]*Japan Society for the Promotion of Science, 5–3–1 Kojimachi, Chiyoda-ku, Tokyo, 102–0083 Japan*

(*Received 13 September 2019, Accepted for publication 27 December 2019*)

Keywords: Korean denasalisation, Nasality, Nasal cavity, Voice onset time
PACS number: 43.71.Hw [doi:10.1250/ast.41.697]

1. Introduction

The coupling of the oral and nasal cavity gives distinctive acoustic properties to speech sounds [1], which in turn bring the sense of nasality to listeners. However, judgement on the presence of nasality is not always straightforward. This increases the difficulty in studying the Korean "denasalisation," that is the losing or weakening of nasality and/or nasalance in phonemic nasal consonants [2]. In Korean, two nasal phonemes /m/ and /n/ appear in the word-initial position, while /ŋ/ appears only in the coda position unless resyllabified. Korean word-initial nasal consonants may be produced with weak or no nasal resonance ([2,3] among others), and even without prevoicing [4]. Such Korean nasals may be perceived as /b/ and /d/ for those who are not native speakers of Korean [3].

Previous studies have demonstrated that Korean word-initial nasals could be heard as oral plosives cross-linguistically using naturally spoken stimuli or spliced speech to test the perception. However, such methods depend heavily on the quality of the original speech, i.e. individual item itself. On the other hand, the "denasalised" Korean nasals do not necessarily lack the nasal energy or nasal airflow, only being weaker than the nasals with proper nasal resonance [2,3]. While nasality can arise from the interaction of a variety of articulatory and acoustic characteristics, the feature or quality of the stimuli that will provoke this perception of nasality has yet to be investigated.

The study focused on how Korean listeners, compared to Japanese listeners, perceive nasality on an oral-nasal continuum with conditions other than nasalance and onset duration being kept constant. (Denasalisation has not been reported in Japanese.) We produced nasal stimuli with a mechanical bent vocal-tract model containing a nasal cavity with a controllable velopharyngeal (VP) port, a "nostril," and a movable lower lip (Fig. 1; see [5] for the details of the vocal tract model). The output sounds yield stable nasalance in accordance with the degree of opening of the VP port. This helps us to directly observe the influence of nasal coupling on perceived nasality.

The current report comprises an identification test, in which native Korean listeners and native Japanese listeners identified the sounds with varying degrees of nasal coupling and onset duration as either [ma] or [ba]. We hypothesised the Korean listeners are more likely to perceive sounds with less nasal resonance and shorter consonant duration as nasal than the Japanese listeners, given the possibly weak nasality and short duration in the production of word-initial nasal onset by Korean speakers. Accordingly, the perceptual boundary was predicted to be located closer to the oral endpoint for Korean than for Japanese listeners.

2. Materials and methods

2.1. Stimuli

The glottal source from a Japanese male speaker pronouncing a sustained /a/ was recorded using a microphone (Sony ECM-MS957) connected to a laptop computer through an audio interface (Edirol UA-25EX) and was then extracted by [PointProcess: To Sound (phonation)] on Praat version 6.0.28 [6].

The extracted glottal source was played through the vocal-tract model (Fig. 1); the source was emitted from a small hole on the metal plate attached to a driver unit (TOA TU-750) which was placed under the model and connected to an audio amplifier (Onkyo MA-500U). While the glottal source was being played, the "lower lip" block of the model was manually manipulated for closure and release of the "oral cavity" as illustrated in Fig. 2. The output signal was again recorded using the microphone, which was set near the "mouth" of the vocal tract model and connected to a recorder

*e-mail: yunjihyeon@gmail.com

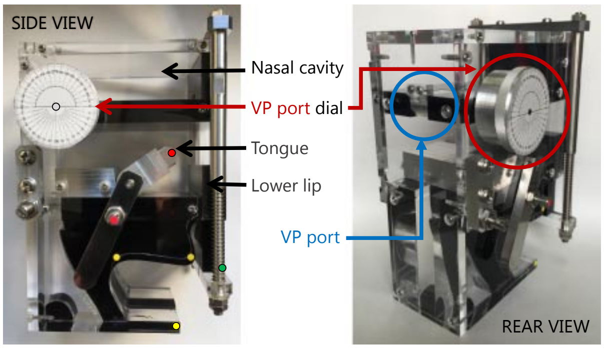

Fig. 1 Mechanical vocal tract model with a nasal cavity.

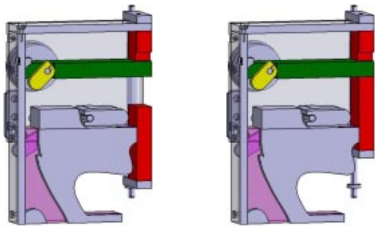

Fig. 2 Release and closure of the mechanical vocal tract model's oral cavity. The VP port is open as the velum (yellow part) is tilted. Images are flipped horizontally for comprehensibility.

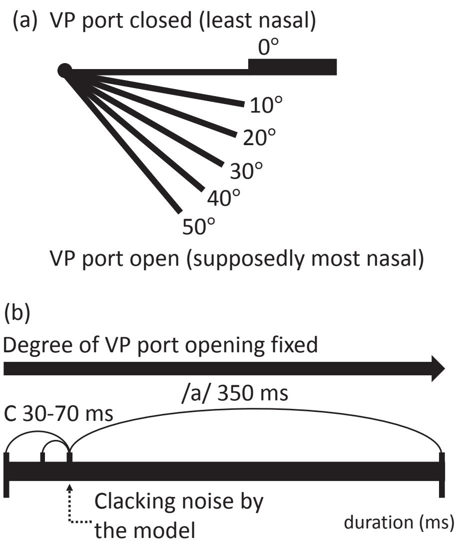

Fig. 3 Schematic illustration of the manipulation of stimuli: (a) VP port opening of 0 to 50 degrees, and (b) onset duration of 30 and 70 ms.

(Marantz PMD661MKII), with a sampling frequency of 48,000 Hz and 24-bit quantisation (mono channel). The manipulation of the sounds is schematically shown in Fig. 3. The VP port was manipulated by rotating the VP-port dial (Fig. 1), opening it in ten-degree steps from 0 (fully closed, that is, no nasal airflow) and 50 degrees (fully open, that is, with the highest nasalance). Rotation of the VP port above

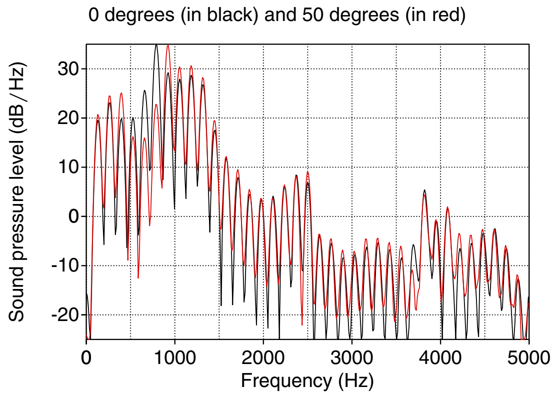

Fig. 4 Power spectra measured at the midpoint of vowel in /Ca/ stimuli with the onset duration of 60 ms: VP-port rotation at 0 and 50 degrees.

50 degrees produces unnatural sounds, so 50 degrees was selected as the most open position for recording. During the operation of the lip movement, the elasticity of the springs caused the lower lip block to hit back against the metal supports at the bottom. Due to this, a clacking noise was introduced to the stimuli a few milliseconds after the release of the "oral" emission.

The recorded samples resembled an /aCa/ sequence. We removed the initial vowel before the intervocalic consonant interval to give an impression of the isolated /Ca/ syllable. Each sound differing in the degree of VP-port opening was edited in Praat [6] to include the onset duration, i.e. the interval before the clacking noise, of 30 to 70 ms in 10 ms steps by removing extra intervals. The duration of vowel interval was 350 ms for all stimuli.

Stimuli with 60 ms onset were acoustically analysed in Praat [6]. The spectra were taken at the midpoint of vowel with a 30 ms Gaussian window. The effect of nasalisation can be observed from Fig. 4, where the power spectrum of the VP-port opening at 0 degrees is shown in black and the spectrum at 50 degrees in red. The change due to nasal coupling is prominent below 1,000 Hz. The nasal coupling raised the intensity of H3 and H7, while lowering the intensity of H4, H5, and H6, compared to the closed VP-port at 0 degrees (Table 1).

2.2. Participants

The experiment was approved by the Research Ethics Committee at Sophia University. A total of 11 native Korean listeners (F4, M7; mean age = 27.73, *sd* = 4.71, *min* = 19, *max* = 35) and 10 native Japanese listeners (F7, M3; mean age = 21.8, *sd* = 1.78, *min* = 20, *max* = 26) participated in the experiment. Among the Korean participants, there were 7 from Seoul/Gyeonggi region, 1 from Gangwon (northeast), 2 from Daejeon (mid-west), and 1 from Gwangju (southwest). Their length of stay in Japan varied between less than 1 month to 12 years (mean: 3.55 yrs, median: 2 yrs). The Japanese listeners were from Chiba/Kanagawa/Saitama/Tokyo region except one participant from Nagoya Aichi (Central region). All listeners reported no deficiency in language and hearing. They gave written informed consent and were reimbursed for their participation.

Table 1 Intensity of the harmonics in vowel with the VP-port closed (vp 00) and with 50 degrees open (vp 50).

		H3	H4	H5	H6	H7
vp 00	frequency (Hz)	395.5	527.3	657.7	790.1	922.8
	intensity (dB)	20.0	20.2	25.2	34.9	29.3
vp 50	frequency (Hz)	394.7	526.7	657.7	789.4	921.3
	intensity (dB)	25.2	16.3	16.0	23.0	34.8

2.3. Procedure

The experiment was conducted in a soundproof room at Sophia University in 2017. It took approximately 20 minutes for each participant. The stimuli were presented through headphones (Sony MDR-CD900ST) via an audio interface (Edirol UA-25EX) which was connected to a laptop computer (Lenovo ThinkPad T410, Windows 8.1). The volume was set to an identical level for all participants at the beginning of the experiment. The participants were allowed to adjust the volume if they felt uncomfortable with the sound level.

The experiment platform was ExperimentMFC in Praat [6]. The instructions were given in Korean to Korean listeners and in Japanese to Japanese listeners, both orally and in written form at the beginning of the session. In each trial, a stimulus was automatically played in random order on a blank screen after 2 seconds of initial silence. Listeners were asked to choose if the sound was /ma/ or /ba/. The options were shown in Katakana—"マ"/ma/ vs. "バ"/ba/ for Japanese listeners, and in Hangeul—"마"/ma/ vs. "바" /pa/ for Korean listeners. In Korean phonology, [ba] is an allophone of /pa/. The order of the two buttons were counter-balanced between participants. Participants could listen to the stimulus of a trial one more time with a "replay" button on the experiment screen, if necessary. They took a self-paced break after every 30 trials. A total of 6,300 data points were obtained (6 VP-port openings $\times$ 5 onset durations $\times$ 10 repetitions $\times$ 21 participants).

3. Statistical analyses and results

Mean responses of /ma/ are shown for each group in Table 2. A generalised mixed-effects model was fitted to the data using the glmer() function of the *lme4* package [7] in R. Fixed effects were degree of VP-port opening (*vp*), duration (*ms*), language group (*lang*), and the two-way interactions of the three predictors. The three-way interaction term was excluded, because a likelihood ratio test found no difference for the models with and without the term [$\chi^2(1) = 1.10$, $p = 0.294$], and the selected model had smaller AIC. Also, by-participant random intercepts were included in the model. For *vp* and *ms*, which were *numeric* predictors, the values were divided by 10 and centred. Thus one unit of *vp* is each 10 degrees of VP-port opening, and one unit of *ms* is each 10 ms onset duration. The model was built on 6,300 observations. The residual deviance of the model is 2549.5 and the residual degrees of freedom is 6292. In the post-hoc tests, slopes of the covariate trends were estimated at their average values based on the fitted model and compared between the language groups using emtrends() of the *emmeans* package in R.

Table 2 Mean proportion of /ma/ responses for 11 Korean listeners (KR) and for 10 Japanese listeners (JP). The 'vp' column indicates the degree of VP-port opening.

		onset duration (in ms)				
group	vp	30	40	50	60	70
KR	0°	0.00	0.01	0.01	0.00	0.01
	10°	0.00	0.04	0.05	0.07	0.05
	20°	0.37	0.70	0.79	0.74	0.67
	30°	0.71	0.90	0.97	0.94	0.88
	40°	0.93	1.00	1.00	1.00	1.00
	50°	0.96	0.99	1.00	1.00	0.99
JP	0°	0.00	0.00	0.00	0.00	0.01
	10°	0.00	0.00	0.03	0.00	0.01
	20°	0.42	0.64	0.77	0.65	0.55
	30°	0.84	0.98	0.96	0.97	0.94
	40°	0.95	1.00	0.99	0.99	0.99
	50°	0.95	0.99	1.00	1.00	1.00

Table 3 Estimated coefficient (β), standard error (*se*), the Wald statistic (z) and p-value for the fixed effects of the generalised mixed-effects model. Intercept reflects the perception of /ma/ by Japanese listeners at the VP-port opening ('vp') of 25 degrees and onset duration ('ms') of 50 ms.

	β	*se*	z	p
(Intercept)	1.67	0.25	6.79	1.15e-11***
vp	3.09	0.13	23.55	2e-16***
ms	0.40	0.06	6.47	9.93e-11***
KR	−0.08	0.34	−0.25	0.807
vp:ms	0.38	0.05	7.12	1.06e-12***
vp:KR	−0.40	0.15	−2.62	0.009**
ms:KR	0.16	0.07	2.18	0.029*

Estimated coefficients of the fixed effects are shown in Table 3. 'KR' indicates Korean. The intercept is the log odds of /ma/ response for Japanese group at the average *vp* and *ms* values, and coefficients are the log odds ratios. While the main effect of KR is not significant ($z = -0.25$, $p = 0.807$), the interaction effects are found for KR with each of *vp* and *ms*. Also, the interaction of *vp* and *ms* is observed [$\beta_{vp:ms} = 0.38$ (± 0.05)].

Estimated marginal mean (cf. [8]) of the /ma/ response is shown at each value of *vp* and *ms* for each group in Fig. 5. The sigmoid identification functions in the upper panels of the figure show that nasal coupling is perceived categorically more or less for both Korean and Japanese. The bottom panels confirm that the effect of nasal coupling overpowers onset duration as there are no vp lines other than 20 degrees to cross 0.5 in probability for both language groups. Meanwhile, the plots show that the Korean group is affected more than the Japanese group by onset duration.

Estimated marginal trends obtained with the emtrends() function can also be confirmed in Table 3. An increase of 10

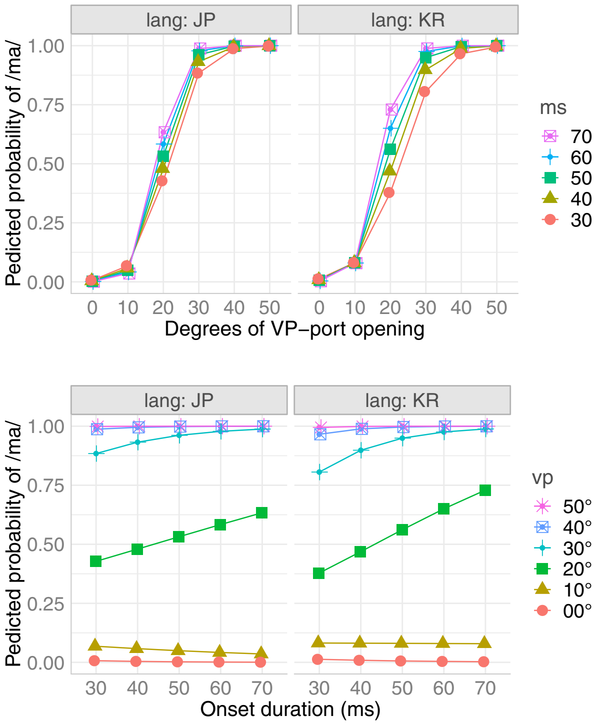

Fig. 5 Predicted probability of /ma/ for the Japanese (JP) and Korean (KR) listeners in terms of VP-port opening (top) and onset duration (bottom).

degrees in the VP-port opening has a greater effect on the perception of /ma/ for both Korean [2.68 ($\pm$0.10), 95% CI: 2.48–2.89] and Japanese listeners [3.09 ($\pm$0.13), 95% CI: 2.83–3.34] compared to the effect of a 10 ms increase in onset duration for Korean [0.56 ($\pm$0.06), 95% CI: 0.45–0.68] and Japanese [0.40 ($\pm$0.06), 95% CI: 0.28–0.52]. JP exhibits a steeper slope for VP-port opening at the 50 ms onset (mean of onset duration) than KR (log odds ratio $= 0.40$, $se = 0.15$, $z = 2.62$, $p = 0.009$) on the logit scale. On the other hand, when *vp* is at its mean, the slope of *ms* is steeper for the Korean listeners (log odds ratio $= 0.16$, $se = 0.07$, $z = 2.18$, $p = 0.029$).

4. Discussion

The current study showed that nasal coupling is more crucial in perceiving /ma/ than onset duration for both Korean and Japanese listeners (Table 3). While short duration is reportedly a characteristic of the denasalised Korean nasals [4], it may not be the duration itself that causes the Korean nasals to be perceived as oral plosive cross-linguistically. The stimuli with greater nasal coupling were consistently perceived as nasal by Japanese listeners regardless of onset duration.

On the other hand, looking at the category boundaries e.g. on the green lines for the *ms* of 50 in the top panels of Fig. 5, there is only slight difference between the Japanese and Korean group. This is somewhat unexpected, given that Korean nasals may lack both nasal resonance and prevoicing [4]. Yet, the magnitude of effects differed for the Korean and Japanese group. The effect of VP-port opening was more gradual for Korean listeners than for Japanese listeners. When the effect of VP-port opening is marginal, Korean listeners had a steeper slope for onset duration than the Japanese listeners. That is, Korean listeners relied on the onset duration more than the Japanese listeners did, when the amount of nasal resonance is ambiguous. Given that a steeper slope in the identification function usually indicates categorical perception, these results imply that while the Korean listeners were not insensitive to nasality, they perceived it less categorically than the Japanese listeners.

The limitation of this study is that the degree of VP-port opening was maintained constant throughout the entire CV syllable of each stimulus. It thus remains unclear whether the perception of nasality was derived from the consonant, the vowel, or both. Further investigation is needed to separate the effect of the consonant and the effect of the vowel. Finally, this study was conducted in Japan, and the Korean participants have been exposed to the Japanese /ba/-/ma/ distinction. It is possible that Korean monolinguals would yield a different result.

Acknowledgements

The contents of this report were presented at the Autumn Meeting of ASJ in 2017. We thank an anonymous reviewer for their suggestions.

References

[1] O. Fujimura, "Analysis of nasal consonants," *J. Acoust. Soc. Am.*, **34**, 1865–1875 (1962).

[2] K. Yoshida, "Phonetic implementation of Korean "denasalisation" and its variation related to prosody," *IULC Work. Pap. Online*, **8**, 1–23 (2008).

[3] Y. S. Kim, *An Acoustic, Aerodynamic and Perceptual Investigation of Word-initial Denasalisation in Korean*, PhD Diss., University College London (2011).

[4] M. J. Ahn, "Acoustic duration of Korean nasals," *Stud. Phon. Phonol. Morphol.*, **19**, 411–431 (2013).

[5] T. Arai, "Integrated mechanical model of [r]-[l] and [b]-[m]-[w] producing consonant cluster [br]," *Proc. Interspeech 2017*, pp. 979–983 (2017).

[6] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," [Computer program], Version 6.0.28., http://www.praat.org/ (2017).

[7] D. Bates, M. Mächler, B. Bolker and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv:1406.5823* (2014).

[8] S. R. Searle, F. M. Speed and G. A. Milliken, "Population marginal means in the linear model: An alternative to least squares means," *Am. Stat.*, **34**, 216–221 (1980).