Perception of synthesized /na/ by Korean listeners

Jihyeon Yun¹,²,* and Takayuki Arai³

¹Graduate School of Science and Technology, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan
²Graduate School of Linguistics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea
³Department of Information and Communication Sciences, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(Received 27 February 2019, Accepted for publication 22 August 2019)

Abstract: Previous research reported that Korean nasal consonants can be denasalized in word-initial position. This study examined the perception of word-initial nasal onset /n/ for native Korean listeners using synthesized /Ca/ stimuli with a Klatt synthesizer. We tested the effects of consonant duration, consonant nasality, and vowel nasalization on perception. In a rating experiment, listeners evaluated the goodness of the stimuli as /na/ on a seven-point scale. The participants generally gave favorable ratings to the stimuli with nasalized vowels. Two-thirds of the participants responded that the stimuli with no nasality are good exemplars of /na/, whereas the other listeners did not. In a yes-no experiment, participants judged if the stimuli were /na/ or not. They responded in similar ways they did in the rating experiment. Many listeners gave positive responses as /na/ even to the stimuli with 0 voice onset time, yet the stimuli with longer prevoicing or nasal murmur were more likely to be perceived as /na/. Vowel nasality affected the perception of /na/, while some listeners preferred oral vowels over the nasalized vowels when they evaluated the /na/-likeness.

Keywords: Korean denasalization, Vowel nasalization, Nasality, Prevoicing, Voice onset time, Klatt synthesizer

PACS number: 43.71.Es [doi:10.1250/ast.41.501]

1. INTRODUCTION

The present study investigated the perception of nasal consonants by native Korean listeners using the synthesized monosyllables /Ca/. Korean has three nasal consonants /m/, /n/, and /N/, and /y/, while /ŋ/ does not appear word-initially. There are no phonemic nasal vowels. Previous research has reported that Korean nasal consonants tend to have a weak or no nasal property in the word-initial position [1–9]. These denasalized or weakened nasals have a short voicing duration or are not voiced at all, and may be accompanied by a release burst [5,6,9].

Genuine nasal sounds are characterized by re-enforced intensity at around 250 Hz, with weak components filling in between 1,000 to 2,500 Hz [10]. However, Korean word-initial nasals may lack voicing as well as nasal formants. In a previous study, out of 288 word-initial nasals, 9% of /m/ instances and 38% of /n/ instances were phonetically voiceless [5]. In Yoshida [2], Korean domain-initial nasal consonants showed weak energy when the acoustic energy from the nostrils was measured with a Nasometer II 6400 (Kay Elemetrics Corp). Nonetheless, nasal energy was present, and it was redefined as “nasal weakening” rather than complete denasalization. Kim YS [4] conducted various types of studies on Korean denasalization, including auditory, acoustic, aerodynamic analyses, and accelerometer measurements. Again, the airflow from the nose was weaker for denasalized Korean nasals than for genuine nasals.

Despite their acoustic properties, there is little ambiguity in perception for native listeners—the denasalized sounds are well perceived as nasal for Korean listeners. All of the Korean oral plosives comprising lax, tense, and aspirated stops are voiceless in the utterance-initial position, thus leaving the phonemic nasal category as the only candidate for having any prevoicing or the shortest voice onset time (VOT). A phonological account has been proposed that Korean denasalization is comparable to the
devoicing of lax stops in the word-initial position [11], in line with the interpretation that Korean lax stops are underlyingly voiced [12].

Consequently, it has been found that when Korean word-medial prevocalic lax consonants are spliced to the word-initial position, they are perceived as nasal by Korean native listeners [4]. Also, Seoul Korean learners of Japanese occasionally pronounced nasal onsets in Japanese words with a weak burst such as [m̩] and [n̩], and they would have more problems distinguishing /m/ and /b/ in perception than in production [13].

It is worth noting that while domain-initial nasal consonants get shorter in both Korean and English, the cross-linguistic difference was found in the subsequent vowel [14]. In the domain initial syllable comprising a nasal onset consonant, a vowel, and a coda consonant (#NVC), the subtraction of the amplitude of the nasal peak around 250–300 Hz from the amplitude of the first formant (A1–P0) indicated that the nasality of the vowel was larger near the onset nasal consonant than at the later time points within the same vowel. In contrast, the vowel nasality was weakest right after the onset nasal consonant and then gradually increased in the same #NVC context in Korean [14]. Given such a difference in production, one possibility with regard to perception is that a native Korean listener would be less affected by the timing and degree of coarticulatory vowel nasalization in judging whether the onset consonant is a nasal, when a native English listener would expect a vowel succeeding a nasal consonant to have more carryover nasalization.

While there have been a few perception studies, they have focused on cross-linguistic perception by non-native listeners [4,15] and learners of the Korean language [7,16], or on native listeners’ lexical access regarding prosodic context [8] using naturally spoken stimuli. These studies have shown the characteristics of Korean denasalization, yet the connection between measured acoustic properties and perceptual judgment has not been systematically tested. There are difficulties in isolating a certain factor from other concomitant cues. We expected synthesized speech allows us to modulate relevant parameters independently.

In this study, we examined the perception of nasal phonemes by Korean native listeners using synthesized monosyllables which resembled either [na] or [da] with several parameter combinations with regard to nasality. Specifically, we tested how Korean listeners perceive nasal and non-nasal consonants with different durations and with or without vowel nasalization. The effects of consonant voicing duration, nasal pole and zero in consonants and vowels, the amplitude of the consonants, the first and second formant and their bandwidths in the vowel following a nasal onset, and different f0 contour were explored. While there are two possible word-initial nasal consonants in Korean phonology, only /na/ was chosen as the nasal target. This was to avoid the possible variation caused by place of articulation which is beyond the focus of the current study. Vowel contexts other than /Ca/ were not used for the same reason. With the synthesized stimuli, we conducted a rating task and a yes-no task to obtain phonetic-level evaluation as well as phonemic-level judgments from Korean listeners.

2. RATING EXPERIMENT

2.1. Methods

2.1.1. Stimuli

The stimuli were synthesized using a cascade-formant synthesizer KLSYN93 Version 2.1 with the XKL interface, which was originally developed by Klatt (see [17,18]). There were 7 sets of stimuli, and each set consisted of a continuum in which the VOT of the consonants ranged from −70 to 0 ms in 10 ms steps. The longest nasal murmur is 70 ms, which represents genuine nasals. For instance, the mean durations of nasal murmur are under 70 ms in three different languages: Korean word-medial [n] in CV.CV context [5] (49 ms), [n] in spontaneous Japanaese (52.8 ms) [19], and nasal onsets in domain initial position (without focus) for native speakers of American English (intonational-phrase initially approx. 50 ms and word-initially approx. 70 ms) [20]. The opposite end of the continuum was 0 VOT, which contained no prevoicing or nasal murmur.

The continua of the stimuli differed in the combination of parameters related to nasal resonance in their consonant and/or vowel interval. The vowel duration was 200 ms for all stimuli, and the entire duration of a stimulus differed according to its VOT. In the synthesis, intervals were updated every 1 ms for parameter reset (UI = 1). The sampling rate was 10,000 Hz.

To synthesize consonants and vowel intervals, the following parameters were manipulated:

**AV** Amplitude of voicing (dB)
**TL** Extra tilt of voicing spectrum (dB down at 3kHz)
**OQ** Open quotient (voicing open time/period, %)
**F0** Fundamental frequency (1/10 Herz)
**B1–B5** Bandwidths of first to fifth formants (Hz)
**A3F–A5F** Amplitude of frication-excited parallel third to fifth formants (dB)
**AI** Amplitude of impulse (dB)
**FNP** Frequency of nasal pole (Hz)
**FNZ** Frequency of nasal zero (Hz)

(Cited from the synthesis specification document provided in XKL)

The input values for synthesis parameters and the names used to refer to each continuum are shown in Appendices (A.1). The prevoicing without nasal resonance
The result suggests the quality of the stimuli are acceptable for the prev continuum. The specific values followed [21] except for TL, for which XKL might have a different scale from Synthworks implementation in the original literature. The burst was inserted at the vowel onset by adding A3F to A5F and A1 only for the stimuli in the prev continuum. For the other 6 continua with nasal murmur (mur, mur_vband, mur_fnsp, nas, nas_f0, nas58), FNP, FNZ, F1 to F3, and B1 to B3 values for [n] before the front vowels in [17] were applied with modification. Additionally, a /Ca/ syllable with an aspiration interval of 75 ms [AH (amplitude of aspiration) = 50 dB] was synthesized and used as a filler in the experiments.

For vowel intervals, the target values for F1 and F2 were by default 700 Hz and 1,200 Hz, respectively. F3 to F5 were kept to the default values of the software. B1 and B2 were 60 Hz and 90 Hz. The conditions for vowel nasalization were included to study coarticulatory effects, and the parameters for nasalized vowels in the mur_fnsp, nas, nas_f0, and nas58 continua were taken per [22]. For the three continua nas, nas_f0, and nas58, the target F1 was modified by increasing both parameters by 100 Hz in relation to FN and FNZ [17,22]. The target F2 was also increased by 100 Hz for these three continua to adjust the vowel quality and amplitude. B1 and B2 were increased by 20 Hz and 10 Hz, respectively, for mur_vband, nas, nas_f0, and nas58 continuum. The amplitude was 60 dB for the initial 150 ms of the vowel, which fell to 0 during the last 50 ms.

The effects of amplitude and f0 contour in the consonant interval were further taken into account. The voicing amplitude (AV) was set to 58 dB for the consonant interval in the nas58 continuum, and to 55 dB for the rest of the continua. The fundamental frequency was fixed at 100 Hz (F0 parameter input value of 1,000 in XKL) for the consonant interval, then rose to 120 Hz during the first 50 ms of the vowel interval and fell to 100 Hz during the last 150 ms by default. For the nas_f0 continuum, f0 started to rise at the consonant-initial, resulting in a slightly higher f0 during the consonant interval. The f0 at the beginning of the vowel was approximately 10 Hz higher for the −70 ms stimulus in the nas_f0 continuum than the stimulus with the same VOT in the nas continuum. The specific values of the parameters are shown in A.1 in Appendices for each continuum.

Twenty native Japanese listeners listened to the prev, mur, and nas continua, and evaluated their overall naturalness as human speech sounds on a 10-point rating scale. The result suggests the quality of the stimuli are acceptable as speech sounds (mean = 6.15, sd = 2.25, median = 7).

2.1.2. Participants

The experiment was approved by the Research Ethics Committee at Sophia University. A total of 23 Korean listeners participated in the rating experiment (17 female, 6 male; age range of 18 to 38 years, mean age of 27.4 years). They were either students or staff members at Chungnam National University in Daejeon, located in the mid-west region of South Korea. Until the age of 15, 4 participants were brought up in the Seoul/Gyeonggi region (hereafter, Gyeonggi), 16 in the Daejeon/Chungcheong region (Chungcheong), 2 in Jeolla, and 2 in Gyeongsang. Individual information is shown in Appendices (A.2). They gave written informed consent, and were reimbursed for their participation. All listeners reported that they have no deficiency in language and hearing.

2.1.3. Procedure

The experiment was conducted in a sound-treated room at Chungnam National University in August 2018. 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 (LG Gram 15ZD980-TX56K, Windows 10). The volume was set to an identical level for all participants at the beginning of the experiment. They could adjust the volume if they felt uncomfortable, but most of the participants did not change the volume. The experiment platform was ExperimentMFC in Praat version 6.0.28 [23].

The instructions were given in Korean both orally and in written form at the beginning of the session. On the start screen, participants were asked to choose the number according to the goodness of the played sound as “na.” Number 0 was explained as a sound that is not good at all as /na/, and number 6 was described as a very good sound as /na/.

Stimuli were presented in random order using the “PermuteBalancedNoDoublets” option. On the experiment screen, the sentence “가 말소리는 [나]일까?” [Is the sound ‘na’?] always appeared on the top. In the middle of the screen, there were 7 buttons, each with one of the numbers from 0 to 6 on them in consecutive order, as well as additional Korean texts “/나/에 [나]” [not /na/] on the left of the “0” button and “/나/ 맛” [it is /na/] on the right of the “6” button. Listeners rated the stimulus between 0 and 6 depending on its “/na/-likeness” by clicking on the button with the corresponding number and pressed an “OK” button to move on to the next trial. They had the option of taking a break after every 30 trials.

A total of 3,864 data points were obtained for the target stimuli (23 participants × 3 repetitions × 8 VOT steps × 7 continua).

2.1.4. Analyses

Statistical analyses were performed in R version 3.5.0. From visual observation, we found that some listeners responded in a way that systematically conflicting with the other listeners (see Fig. 1 in Sect. 2.2 for the variation).
Since the actual trends in the data could be canceled out in arithmetic means if pooled, there was a need to capture this between-participant and between-continuum variation.

As a means to inspect the similarity among the continua, a principal component analysis (PCA) was performed on the responses that were scaled and aggregated by continua for each participant, using the `prcomp()` function. Meanwhile, the listeners were divided into two groups according to how they responded to the prev continuum. PrevHigh group consisted of the listeners who gave mean responses of greater than 3 to prev. The rest of the listeners belonged to PrevLow group.

In further statistical analyses, only the responses to prev, mur, and nas continua were taken into account; the data for the rest of the continua were excluded for simplification. Also, only 19 participants who were originally from Gyeonggi or Chungcheong were included to reduce the possible influence of dialects.

We fitted linear mixed-effects models separately for each group with the `lmer()` function in the `lmerTest` package. For model selection, the models were fitted by maximum likelihood with the Satterthwaite approximation to the degrees of freedom for t-tests. In the models, the obtained /na/-likeness points were the response variable. VOT step (absolute values of VOT divided by 10, a numeric variable) were added as a fixed effect term together with types of continua (a factor variable) and the interaction of the two. For random effects, by-participant random intercepts were included. This would help to reduce the possible influence of dialects.

We compared each of the models above to another model with the same variables but without an interaction term. If a likelihood ratio test using `anova()` finds no significant difference between the models at 95% confidence level, we chose the model without an interaction term and checked whether the interaction term improves model fit. If a likelihood ratio test using `anova()` finds no significant difference between the models at 95% confidence level, we chose the model without an interaction term. Otherwise, the model with a smaller AIC was selected, and a post-hoc test was carried on; the model was passed to the `emtrends()` function in the `emmeans` package for post-hoc pairwise comparison with the Satterthwaite approximation to estimate degrees of freedom and with the Tukey method for p-value adjustment.

### 2.2. Results

Figure 1 shows each participant’s mean responses for each continuum. There was a huge variability among listeners, but a certain pattern was observed. For instance, participant K07, shown on the topmost row in Fig. 1, gave higher points to the prev, mur, and mur_vband stimuli, while giving comparatively lower points to mur_fnp, nas, nas_f0, and nas58. In contrast, K19 on the bottom of Fig. 1 preferred the latter four continua to the former prev, mur, and mur_vband as /na/. All 69 trials with the filler received 0 points.

2.2.1. Principal component analysis on continua

The result of the PCA suggests perceptual similarities among the continua. When between-participant variations were given as factors, PCA yielded 7 dimensions of principal components. The first (73.68%) and second (18.02%) principal component cumulatively explained 91.70% of the listener variance in the continua (Fig. 2).
Based on the proximities on the first two dimensions, the continua were grouped into 3 clusters as below:  
**Oral vowel continua:** prev, mur, mur_vband  
**Nasal vowel continua:** nas, nas58, nas_f0  
**Isolated:** mur_fnp  

### 2.2.2. Mean responses  

The participants were divided into two groups depending on their response to the prev continuum. The members of each group are indicated in Appendices (A.2). The four listeners originally from Jeolla or Gyeongsang were excluded from further statistical analyses. The responses were then aggregated over the 19 participant either from Gyeonggi or Chungcheong, by VOT and continuum for each of the PrevHigh (n = 13) and PrevLow (n = 6) group (Fig. 3).

The PrevHigh group gave the highest scores to prev and the other oral vowel continua. The nasal vowel continua were given the second highest points, and mur_fnp received the lowest score. Contrastively, the PrevLow group evaluated nasal vowel continua (nas, nas58, and nas_f0) as the best /na/ sounds, while rating the oral vowel continua as the lowest; mur_fnp, for which the nasal pole and zero were synthesized in the vowel interval without explicit adjustment of F1 and F2 input, was between the other two continua clusters. Table 1 shows mean responses for each continuum. The mean /na/-likeness of the prev continuum was the highest for the PrevHigh group (mean = 5.09), but the lowest for the PrevLow group (mean = 1.41).

#### 2.2.3. Linear mixed-effects models  

Linear mixed-effects models provide an intercept and different slopes for VOT steps and types of continua. In fitting the models for the PrevHigh group, 936 observations from 13 participants were used. For PrevLow, 432 data points from 6 listeners were used. Likelihood ratio tests suggested no interaction between VOT step and continuum for the PrevHigh listeners [$\chi^2(2) = 2.31, p = 0.316$] while confirming the interaction effect for the PrevLow listeners [$\chi^2(2) = 12.171, p = 0.002$]. The interaction term was thus added to the PrevLow model only.

Fixed effects of the PrevHigh and the PrevLow models are shown in Table 2. Here, the intercepts reflect the prev stimulus with 0 VOT. In the PrevHigh model, VOT step, mur, and nas had an effect that differs from the intercept. A single VOT step raised /na/-likeness point by 0.09. Compared with prev, the mur and nas continua reduced the /na/-likeness points by −0.45 and −0.72, respectively. Predicted responses for /na/ with 95% confidence intervals are separately shown for each listener group in Figs. 4 and 5. The responses predicted from the PrevHigh model (Fig. 4) are generally located between the /na/-likeness scores of 4 and 6. This is the case even at 0 VOT, that is, when there is no prevoicing. This is clearly different from the PrevLow model (Fig. 5). The predicted responses at 0 VOT are below 2 for both prev and mur, which are much lower than the response for nas.

The PrevLow model suggests a significant main effect of the continuum nas and the interaction of the VOT step and the continuum mur (Table 2). Estimated marginal trends for the three continua and the result of pairwise

---

**Table 1** PrevHigh and PrevLow listeners’ mean responses of /na/-likeness are averaged over VOT step and shown for each continuum. Fractional amounts are rounded off to the second decimal place.

| continuum | PrevHigh | PrevLow |
|-----------|----------|---------|
| prev      | 5.09     | 1.41    |
| mur       | 4.64     | 2.17    |
| mur_vband | 4.52     | 2.31    |
| mur_fnp   | 3.67     | 3.25    |
| nas       | 4.37     | 4.74    |
| nas_f0    | 4.53     | 4.62    |
| nas58     | 4.45     | 4.81    |

---

**Fig. 2** Seven continua shown on the first (Dim1) and the second (Dim2) dimensions obtained from the principal component analysis.

**Fig. 3** Group means of responses for each continuum on the voice onset time from 13 PrevHigh listeners and 6 PrevLow listeners.
### 2.3. Discussion

The PCA results show that prev, mur, and mur_vband clustered together, with mur and mur_vband being located closer to each other. Notably, the oral vowel continua induced responses that are similar to one another from each participant, regardless of whether the listener gave them high or low points. The PrevLow group who preferred nasalized vowels gave lower scores equally to mur and mur_vband with a nasal consonant and to prev containing an oral prevocalic consonant. The vowel nasalization thus affected the perceptual judgment more than the consonant did, while whether the property was taken positively or not varied among the participants.

There was a group of listeners who perceived the continua with an oral vowel as /na/, or even as better than the continua with a nasal consonant and a nasalized vowel (Fig. 4). The fact that they judged the stimuli with no nasal resonance as the best representations for /na/ over the other nasal stimuli may be a characteristic of Korean listeners. Even the stimuli with no prevoking or nasal murmur obtained positive responses as /na/.

#### Table 3

Pairwise comparisons among the three continua prev, mur, and nas in the rating experiment. Estimated marginal trend (trend), standard error (se), and lower and upper limit of 95% confidence interval (LCL and UCL) for each continuum at −35 ms VOT step are shown in the top Table. Pairwise comparisons are shown on the bottom; the difference in estimated marginal trends between two continua is shown in ‘estimate’ column with standard error (se), t-ratio and p-value.

| continuum | trend | se  | LCL  | UCL  |
|-----------|-------|-----|------|------|
| prev      | −0.03 | 0.05| −0.13| 0.07 |
| mur       | 0.22  | 0.05| 0.12 | 0.32 |
| nas       | 0.05  | 0.05| −0.05| 0.15 |

| contrast   | estimate | se  | t ratio | p    |
|------------|----------|-----|---------|------|
| prev - mur | −0.25    | 0.07| −3.43   | 0.002|
| prev - nas | −0.08    | 0.07| −1.05   | 0.546|
| mur - nas  | 0.17     | 0.07| 2.38    | 0.047|

#### Fig. 5

Predicted /na/-likeness response from PrevLow group listeners’ model in the rating experiment. A regression line for nas appears on the top, for mur in the middle, and for prev on the bottom at −35 ms VOT.

#### Table 2

Estimated coefficient ($\beta$), standard error (se), t- and p-value of the fixed effects of linear mixed-effects models for PrevHigh and PrevLow groups in the rating experiment. Fractional amounts are rounded off to the second decimal place for $\beta$, se, and t-value. Three significant figures are shown for p-value.

| group       | $\beta$ | se  | t    | p     |
|-------------|---------|-----|------|-------|
| PrevHigh    | (Intercept) | 4.76| 0.23 | 21.11| 5.31e-16*** |
|             | VOTstep | 0.09| 0.02 | 4.20 | 2.96e-05*** |
|             | mur     | −0.45| 0.13| −3.56| 3.84e-04*** |
|             | nas     | −0.72| 0.13| −5.71| 1.50e-08*** |
| PrevLow     | (Intercept) | 1.51| 0.25 | 5.94 | 3.19e-07*** |
|             | VOTstep | −0.03| 0.05| −0.55| 0.582 |
|             | mur     | −0.12| 0.31| −0.39| 0.694 |
|             | nas     | 3.06 | 0.31| 10.02| <2e-16*** |
|             | VOTstep:mur | 0.25| 0.07| 3.43 | 6.65e-04*** |
|             | VOTstep:nas | 0.08| 0.07| 1.05 | 0.295 |

(***p < 0.001)
The cases are different for the PrevLow listeners as shown in Fig. 5. They gave higher points to nas throughout the VOT scale, while giving much lower points to the 0 VOT stimuli of the prev and mur. Then, for mur with a nasal consonant, the /na/-likeness score increased as the VOT step increases, whereas the /na/-likeness points gradually decreased for prev with an oral prevocalic consonant. The /na/-likeness scores for the nasal vowel continua were generally high for both PrevHigh and PrevLow.

The mur_fnp continuum was rather distinct from the other continua and constituted a separate cluster in the PCA. mur_fnp is a continuum with a nasal consonant and nasalized vowel. What make it different from nas are the F1, F2, B1, and B2 input values for the vowel interval; those were kept the same as the oral vowel continua for mur_fnp, whereas for nas, both F1 and F2 were explicitly raised by 100 Hz with slightly broader B1 and B2. The intensity of the vowel was weaker in mur_fnp due to the nasal pole and zero pair. Also, the nasal pole and zero changed the spectrum and made the vowel closer to another Korean vowel /o/. The other nasal vowel continua did not have the same problem by virtue of the explicit adjustment of vowel formants. While the PrevHigh group gave the lowest score to mur_fnp, the PrevLow group gave it intermediate points that were between the nasal vowel continua and the oral vowel continua. The PrevLow listeners have given a lower score to the mur_fnp continuum presumably due to the less-prototypical vowel quality, yet they evaluated this continuum higher than the oral vowel continua for the nasal resonance.

The effects of the modifications in the f0 contour (nas_f0) and the amplitude (nas58) were subtle, possibly due to the ceiling effect. Similarly, the increase of 20 Hz for B1 and 10 Hz for B2 did not make noticeable differences on the responses for mur_vband from mur. We infer that the degree of modification was too small.

3. YES-NO EXPERIMENT

In a rating experiment, listeners evaluate stimuli subjectively, and the responses may reflect fine-grained quality of the stimuli. However, categorical judgments with forced choices may yield different results. If Korean listeners perceive the stimuli phonologically, it is possible that they respond the entire stimulus sets are /na/, which implies the phonetic details of nasal resonance does not affect their phonemic perception. In order to look into this possibility, we conducted another experiment in which listeners were asked to judge whether each stimulus was /na/ or not. A yes-no paradigm was chosen over a labeling task, because no typical oral consonant in Korean is produced with a negative VOT.

3.1. Participants

Fourteen Korean native listeners, ranging in age from 22 to 38 (mean 27.3), participated in a yes-no experiment (10 female and 4 male). All 14 listeners also participated in the rating test (c.f. A.2 in Appendices). There were 4 participants originally from Gyeonggi, 8 from Chungcheong, one from Jeolla and one from Gyeongsang.

3.2. Procedure

The test conditions were identical to the rating experiment with regard to stimuli, filler with 75 ms voicing lag, equipment, software, and experiment location, as described in Sects. 2.1.1 and 2.1.3.

Initially, the instruction was given in Korean to each participant. They were asked to respond “yes” or “no” to the question “is it /na/?” by clicking on corresponding buttons. In each trial, a stimulus was automatically played in random order after 2 seconds of initial silence. Then a sentence “이 말소리는 /나/입니까?” [is the sound ‘na’?] appeared on the top of the experiment screen. For the yes-no test, a “ye” [yes] button was placed on the left-hand side of the experiment screen, and a “no” button was on the right-hand side. The labels and messages were written in Hangeul, the Korean writing system. The participant could click a “Replay” button on the bottom of the screen to listen to it one more time, or use a “Back” button on the bottom-left side to go back to the previous trial. Participants were encouraged to use the “Replay” or “Back” button as little as possible, unless they lost attention, could not hear the sound at all, or mistakenly clicked on the wrong button.

They could take a self-paced break after every 30 trials. Each stimulus was presented 4 times in a session. A total of 3,192 data points was obtained (57 stimuli × 4 repetitions × 14 participants).

3.3. Analysis

To summarize the results, we first divided the listeners into two groups based on mean responses as /na/ for the prev continuum. PrevHigh group in the yes-no experiment was composed of listeners who had mean responses of greater than 0.5 for prev. PrevLow group listeners were those with the mean responses of less than or equal to 0.5 for prev. The data of the two participants from the southern region were excluded before further analyses.

For each group, a generalized mixed effects model was fitted with a logit link function using glmer() of the lme4 package in R. The “yes” and “no” responses were dummy-coded as 1 and 0, respectively, and they were added as a response variable. We used only the responses on the prev, mur, and nas continua for the sake of simplicity. For fixed effects, VOT step (absolute VOT values divided by 10, a numeric variable) ranging from 0 to 7 were added together.
with type of continuum (a factor variable) and their interaction. By-participant random intercepts were also specified. When the model with an interaction term is chosen based on a likelihood ratio test, a post-hoc test was run on the fitted model, in which pairwise comparisons of estimated slopes were carried out for the continua with the Tukey adjustment, using `emtrends()` of the `emmeans` package in R.

### 3.4. Results

The listeners were divided into two groups depending on whether the proportion of their mean responses on the prev stimuli was greater than 0.5 (PrevHigh, \( n = 9 \)) or not (PrevLow, \( n = 5 \)). Although this grouping criterion was independent of the grouping in the rating experiment, all of the listeners who were categorized as PrevHigh in the yes-no experiment also belonged to the PrevHigh group in the rating experiment, and all PrevLow listeners in the yes-no experiment were the members of the equivalent group in the rating experiment.

#### 3.4.1. Mean responses

The mean proportion of /na/ is shown for PrevHigh listeners (\( n = 8 \)) and PrevLow listeners (\( n = 4 \)) separately in Fig. 6. The mean proportion of /na/ averaged over VOT step for each continuum are shown in Table 4. All 56 trials with the filler containing aspiration obtained “no” responses.

#### 3.4.2. Generalized mixed-effects model

When the data were divided into the PrevHigh and PrevLow group, the generalized mixed-effects model for the PrevLow group did not converge. Instead of grouping, we pooled the 12 listeners’ data and fitted a generalized mixed-effects model with 1,152 data points. The fixed effects of this model is shown in Table 5.

#### Table 4  Mean proportion of /na/ responses for PrevHigh and PrevLow groups in the yes-no experiment.

| Continuum   | PrevHigh | PrevLow |
|-------------|----------|---------|
| prev        | 0.88     | 0.20    |
| mur         | 0.89     | 0.46    |
| mur_vband   | 0.91     | 0.91    |
| mur_fnp     | 0.78     | 0.55    |
| nas         | 0.89     | 1.00    |
| nas_f0      | 0.89     | 0.99    |
| nas58       | 0.87     | 0.99    |

#### Table 5  Estimated coefficient (\( \beta \)), standard error (se), \( z \) and \( p \)-value for the fixed effects of the generalized mixed effects-model in the yes-no experiment. Coefficients are shown in odds ratios.

| Pooled data (12 listeners) | \( \beta \) | se | \( z \) | \( p \) |
|----------------------------|------------|----|--------|--------|
| (Intercept)                | 0.83       | 0.64| 1.30   | 0.193  |
| VOTstep                    | 0.11       | 0.06| 2.02   | 0.043* |
| mur                        | 0.16       | 0.34| 0.48   | 0.634  |
| nas                        | 2.63       | 0.44| 6.02   | 1.79e-09*** |
| VOTstep:mur                | 0.15       | 0.08| 1.74   | 0.082  |
| VOTstep:nas                | −0.08      | 0.16| −0.76  | 0.450  |

#### Fig. 6  Mean proportion of “yes” (/na/) responses for PrevHigh and PrevLow group listeners in the yes-no experiment.

#### Fig. 7  Predicted probability of /na/ based on 12 listeners’ data: nas appears on the top, mur in the middle, and prev on the bottom at −35 ms VOT.

In the model, the residual deviance is 875.0 and the residual degrees of freedom is 1,145. A likelihood ratio test suggested that the model with VOT step, continuum and their interaction had a better fit than its cognate model without an interaction term at the 90% confidence level \( \chi^2(2) = 5.23, p = 0.073 \). Figure 7 is predicted probability of /na/ by the continua based on the model estimation. In Table 5, the intercept represents the stimuli with 0 VOT continuum. The nas continuum was different from this intercept \( \beta = 2.63 (\pm 0.44), z = 6.02 \). A small
3.5. Discussion

Similar to the rating test, there were between-participant variability in how they responded to the oral vowels. The PrevHigh group in the yes-no experiment, again, tended to judge all of the stimulus sets as /na/, giving mean proportions that were higher than 0.6. The difference between the oral vowel continua and the nasal vowel continua was small for this group (Fig. 6). The responses of the PrevLow listeners in the yes-no experiment were also similar to the results for the equivalent group in the rating experiment. In fact, all of the PrevLow listeners in the yes-no experiment were also members of the PrevLow group in the rating experiment. They judged the stimuli with a nasalized vowel as better /na/, and the stimuli with a prevocalic consonant to be poor. The nas, nas_f0, and nas58 were the highest, given the mean proportions between 0.9 and 1. Prev received the least /na/ responses, and the other three continua with nasal-related properties were in the middle. The mur and mur_yband stimuli showed a stronger effect of VOT (Fig. 6 and Table 6). That is, duration of voicing in the consonant interval had a stronger effect on the PrevLow listeners’ perception when stimuli had a nasal consonant but no carryover vowel nasalization than for the other conditions. This shows that the PrevLow group is judging based on the nasal resonance. However, which factors explain the variability between the PrevHigh and PrevLow groups is not clarified.

Table 6 Estimated marginal trend on the logit scale, standard error (se), and lower and upper limit of 95% confidence interval (LCL and UCL) for each continuum at −35 ms VOT (top), and pairwise comparison of estimated marginal trends (bottom).

| continuum | trend | se | LCL | UCL |
|-----------|-------|----|-----|-----|
| prev      | 0.12  | 0.06 | 0.00 | 0.23 |
| mur       | 0.26  | 0.06 | 0.14 | 0.38 |
| nas       | 0.04  | 0.09 | −0.14 | 0.21 |

| contrast | estimate | se | z ratio | p     |
|----------|----------|----|---------|-------|
| prev - mur | −0.15 | 0.08 | −1.74 | 0.191 |
| prev - nas | 0.08 | 0.16 | 0.76 | 0.730 |
| mur - nas | 0.23 | 0.11 | 2.08 | 0.093 |

The effect of the interaction between VOT step and the mur continuum was found at the 90% confidence level [β = 0.15 (±0.08), z = 1.74].

Table 6 is the result of the post-hoc test. A difference in odds ratio was observed between the slopes of mur and nas at 90% confidence level. The predicted probability for prev was not significantly different from mur as well as from nas.

4. GENERAL DISCUSSION AND CONCLUSION

Word-initial Korean nasals are known for very short or no voicing duration. Yet, in this study, longer prevocing or nasal murmur in consonants raised the frequency of nasal perception for Korean listeners. The stimuli with nasal consonant and nasalized vowel tended to be perceived as /na/ for all listeners. Those stimuli were perceived as nasal across the entire range of the consonant duration including 0 ms, thus it is evident that the critical cue was in the vowel.

On the other hand, two thirds of the participants judged oral prevocalic consonants as good nasals. This result is in accordance with the characteristics of Korean denasalization reported in previous research [4]. In the rating experiment, these listeners preferred oral vowels to nasalized vowels, while more or less accepting the entire stimulus sets as /na/. They were even more likely to judge all of the stimuli as /na/ in the yes-no experiment. The previous study reported that Korean and English differed in vowel nasalization: for Korean, there was a stronger effect of nasalization in the vowel succeeding a domain-initial nasal consonant [14]. This explains the perception for these PrevHigh listeners of the present study, who less preferred nasalized vowel while perceiving them as nasal.

In contrast, one third of the listeners in the rating experiment gave higher scores to the stimuli in which both the consonant and vowel were nasal than to the stimuli without nasal resonance. They judged not only the stimuli with prevocalic consonants as poor nasals, but were also likely to give lower scores to /na/ without vowel nasalization.

Looking at the vowel nasalization, many listeners accepted the nasalized vowels only when their F1 and F2 were explicitly adjusted, which made the vowel quality lower and more fronted. When the vowel formants were not adjusted by the synthesis parameters of F1 and F2, the combination of nasal consonant and nasal vowel (mur_fnp) received much lower scores. Synthesizing nasal pole and zero with FNP and FfKZ changes the vowel’s spectral characteristics. We assume that this modification in the spectral envelope affected the perception at phonemic and sub-phonemic levels. When the vowel formants were measured in Praat [23] for the −70 ms stimuli, F1 and F2 were 715 Hz and 1,213 Hz for prev, 520 Hz and 1,173 Hz for mur_fnp, and 600 Hz and 1,272 Hz for nas. The measured formants are shown in Fig. 8 together with the mean F1 and F2 for male speakers of Korean from previous research.

According to an acoustic analysis of read speech, the mean values for F1 and F2 for 5 male speakers of Seoul Korean were 761 Hz (±43) and 1,224 Hz (±61) for /a/,
and F1 = 544 Hz (±23) and F2 = 943 Hz (±65) for the rounded mid-back vowel /a/ [24]. Also, in the speech corpus of read-style Korean, the mean values for F1 and F2 for the analyzed male speaker data were 638 Hz (±38) and 1,333 Hz (±105) for /a/, and 486 Hz (±42) and 1,100 Hz (±64) for /æ/ [25]. The cause of the lower scores for mur_fnp could be that the measured formants of mur_fnp are shifting towards another Korean vowel /æ/. Also, the overall intensity of the vowel in mur_fnp was softer than other continuia, due to the nasal pole and zero.

Although the mur_vband continuum had broader bandwidths than mur, the stimuli with a nasal consonant and an oral vowel, the modification of 20 Hz for B1 and 10 Hz for B2 were not sufficient to affect the perception. The effects of the higher amplitude and the early f0 rising in the consonant interval were little as well, given that the nasalized vowel alone already had a very strong effect.

Korean listeners from different dialect regions participated in this study, but we used only the data from Seoul/ Gyeonggi and Daejeon/Chungcheong listeners when analyzing pooled means and regressions in order to prevent possible variation. Four out of 23 listeners in the rating experiment were originally from the southern region. Given Yoshida (2008), if the dialects affected the responses, the prediction would be that the four listeners from the southern region are less likely to denasalize or to devoice word-initial nasals. In that case, these listeners could have given lower scores to the oral prevocalic stimuli than the listeners from the northern and central regions did. The two listeners from Gyeongsang (southeastern) were members of the PrevLow group, however, the two listeners from Jeolla (southwestern) were PrevHigh, the group who gave high scores to the oral stimuli. In addition, the other members of the PrevLow group were either from Gyeonggi (n = 1) or Chungcheong (n = 2).

To test the effect of dialects, we fitted a linear mixed-effects model with the data from the rating experiment. VOT step, type of continuum, dialect and their three-way interaction were set as fixed effects. By-participant random intercepts were added as well. This model showed that Gyeonggi was not different from Chungcheong (t = 0.21, p = 0.831), but Gyeongsang was (t = 2.73, p = 0.011), and an interaction of Jeolla and continua was found. A post-hoc test found differences in slopes for some dialect contrasts. For prev, differences between the slopes for Chungcheong and Jeolla (estimate = −0.39, se = 0.14, t = −2.89, p = 0.026) and between Gyeonggi and Jeolla (estimate = −0.42, se = 0.16, t = −2.66, p = 0.047) were found. The slope was steeper for Jeolla than for Chungcheong and Gyeonggi, hence the two Jeolla listeners were more affected by duration for the stimuli with the oral prevocalic consonant. For mur and nas, no difference was found for any pair of dialects. This result might suggest that Jeolla and Gyeongsang listeners were different from Gyeonggi and Chungcheong listeners. In any event, interpretation of the approximations on dialect would not be reliable for the current data, given the imbalance and small number of participants compared to the huge variability. Yoo KY [6] found that the older Busan (Gyeongsang region) speakers did not denasalize, while the younger age group showed denasalization with between- and within-speaker variation. Since the participants of the present study were of younger generation, they also might have had such variation.

This study investigated the effects of consonant duration, prevocing, nasal murmur, and vowel nasalization on Korean listeners’ perception of word-initial nasal consonants. One implication is that even if the synthesis parameters are manipulated in linear steps, the effect would not be as linear, not only on listeners perception but also on the acoustic outputs. Based on these results, more balanced tests should be carried out in order to control the variance and to reduce the ambiguity that occurred in the present experiment. How native listeners of other languages perceive the stimuli needs further elucidation.

ACKNOWLEDGMENTS

The contents of this study were presented at New Sounds 2019. We are grateful to our participants, Prof. Cheoljae Seong and the members of the Department of Linguistics at Chungnam National University, Dr. Kanako Tomaru, Dr. Jiayin Gao, Dr. Justine Hui, and two anonymous reviewers.
REFERENCES

[1] 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.

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

[3] S. J. Kim and H. S. Kim, “Aerodynamic approaches to Korean nasals: Phonetic variations focusing on denasalization,” Eoneo Gwahak, 15, 186–193 (2008).

[4] Y. S. Kim, An Acoustic, Aerodynamic and Perceptual Investigation of Word-initial Denasalization in Korean, PhD diss., University College London (2011).

[5] M. J. Ahn, “Acoustic duration of Korean nasals,” Stud. Phon. 270, 460

[6] K. Yoo, “Domain-initial denasalization in Busan Korean: A cross-generational case study,” Proc. ICPhSXVIII, pp. 0845.1–5 (2015).

[7] J. Yoo and S. Kang, “The effect of L2 experience on perception of Korean nasals,” Phon. Speech Sci., 8(4), pp. 63–69 (2016).

[8] S. Shin and A. Tremblay, “Effect of prosodic context on lexical access: An investigation of Korean denasalization,” Proc. 9th Int. Conf. on Speech Prosody, pp. 408–412 (2018).

[9] H. Kim, “Linguistic and social dimensions of denasalization in Seoul Korean,” Proc. HISPhonCog, pp. 106–107 (2018).

[10] S. Hattori, K. Yamamoto and O. Fujimura, “Nasalization of vowels in relation to nasals,” J. Acoust. Soc. Am., 30, 267–274 (1958).

[11] T. Y. Kim, Devicing and Denasalization in Korean: Reconsideration of Phonological Structure in Korean, MA thesis, Waseda University (2015).

[12] M. R. Kim and S. Duanmu, “‘Tense’ and ‘lax’ stops in Korean,” J. East Asian Linguist. 13, 59–104 (2004).

[13] H. Matsuzaki, “Phonetic education of Japanese for Korean speakers,” J. Phon. Soc. Jpn., 3(3), pp. 26–35 (1999).

[14] J. Jang, S. Kim and T. Cho, “Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences,” J. Acoust. Soc. Am., 144, EL33–EL39 (2018).

[15] J. Yun and T. Arai, “Perception of Korean nasal onset /m/ by Japanese listeners: A preliminary study,” Acoust. Sci. & Tech., 39, 424–427 (2018).

[16] E. Yoon and H. Lee, “A perception study on Korean denasalization,” Biling. Res. (J. Korean Soc. Biling.), 55, 183–216 (2014).

[17] D. H. Klatt, “Software for a cascade/parallel formant synthesizer,” J. Acoust. Soc. Am., 67, 971–995 (1980).

[18] D. H. Klatt and L. C. Klatt, “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” J. Acoust. Soc. Am., 87, 820–857 (1990).

[19] T. Arai, N. Warner and S. Greenberg, “Analysis of spontaneous Japanese in a multi-language telephone-speech corpus,” Acoust. Sci. & Tech., 28, 46–48 (2007).

[20] T. Cho, D. Kim and S. Kim, “Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English,” J. Phon., 64, 71–89 (2017).

[21] J. R. Benkı´, “Perception of VOT and first formant onset by Spanish and English speakers,” Proc. 4th Int. Symp. Bilingualism, pp. 240–248 (2005).

[22] S. Hawkins and K. N. Stevens, “Acoustic and perceptual correlates of the non-nasal–nasal distinction for vowels,” J. Acoust. Soc. Am., 77, 1560–1575 (1985).

[23] P. Boersma and D. Weenink, “Praat: Doing phonetics by computer,” [Computer program], Version 6.0.28., http://www.praat.org/ (2017).

[24] C.-J. Seong, “An acoustic analysis on the Korean 8 monophthongs: With respect to the acoustic variables on the F1/F2 vowel space,” J. Acoust. Soc. Korea, 23, 454–461 (2004).

[25] T.-J. Yoon and Y. Kang, “Monophthong analysis on a large-scale speech corpus of read-style Korean,” Phon. Speech Sci., 6, 139–145 (2014).

APPENDICES

A.1 Parameters values. Stimulus-default values are in black and modified values are in blue and bold. The values are shown in dB for AV, and in Hz for the other parameters.

| C | f0 | V |
|---|---|---|
| FNP/FNZ | AV | FNP/FNZ | F1/F2 | B1/B2 |
| prev | 280, 280 | 55 | Vowel onset | 280, 280 | 700, 1,200 | 60, 90 |
| mur | 270, 460 | 55 | Vowel onset | 280, 280 | 700, 1,200 | 60, 90 |
| mur_vband | 270, 460 | 55 | Vowel onset | 280, 280 | 700, 1,200 | 80, 100 |
| mur_fnp | 270, 460 | 55 | Vowel onset | 400, 600 | 700, 1,200 | 60, 90 |
| nas | 270, 460 | 55 | Vowel onset | 400, 600 | 800, 1,300 | 80, 100 |
| nas58 | 270, 460 | 58 | Vowel onset | 400, 600 | 800, 1,300 | 80, 100 |
| nas_f0 | 270, 460 | 55 | Consonant onset | 400, 600 | 800, 1,300 | 80, 100 |
A.2 Participant information. Area indicates the region of residence before age 15. PrevHigh group is the listeners with mean /na/-likeness score (ranging between 0 and 6) of greater than 3 for prev stimuli in the rating experiment and with mean proportion of greater than 0.5 for prev stimuli in the yes-no experiment. PrevLow group comprises the rest of the participants. Seoul is included in Gyeonggi, and Daejeon is included in Chungcheong.

| ID  | Age | Area    | Group     | Rating | Yes-no |
|-----|-----|---------|-----------|--------|--------|
| K01 | 18  | Jeolla  | PrevHigh  | —      | —      |
| K02 | 24  | Chungcheong | PrevLow | PrevLow | —      |
| K03 | 29  | Gyeonggi | PrevLow  | PrevLow | —      |
| K04 | 27  | Chungcheong | PrevHigh | PrevHigh | —      |
| K05 | 36  | Gyeongsang | PrevLow | PrevLow | —      |
| K06 | 26  | Jeolla  | PrevHigh  | PrevHigh | —      |
| K07 | 34  | Chungcheong | PrevHigh | PrevHigh | —      |
| K08 | 22  | Chungcheong | PrevHigh | PrevHigh | —      |
| K09 | 22  | Chungcheong | PrevHigh | PrevHigh | —      |
| K10 | 25  | Gyeonggi | PrevHigh  | PrevHigh | —      |
| K11 | 28  | Chungcheong | PrevLow | PrevLow | —      |
| K12 | 30  | Chungcheong | PrevHigh | —      | —      |
| K13 | 37  | Gyeongsang | PrevLow | —      | —      |
| K14 | 34  | Chungcheong | PrevHigh | —      | —      |
| K15 | 22  | Chungcheong | PrevLow | —      | —      |
| K16 | 29  | Chungcheong | PrevHigh | —      | —      |
| K17 | 31  | Chungcheong | PrevHigh | —      | —      |
| K18 | 23  | Chungcheong | PrevLow | PrevLow | —      |
| K19 | 21  | Chungcheong | PrevLow | —      | —      |
| K20 | 26  | Chungcheong | PrevHigh | —      | —      |
| K21 | 38  | Gyeonggi | PrevHigh  | PrevHigh | —      |
| K22 | 24  | Gyeonggi | PrevHigh  | PrevHigh | —      |
| K23 | 24  | Chungcheong | PrevHigh | PrevHigh | —      |

Total number of participants 23 14