Acoustic Cues of Vowel Quality to Coda Nasal Perception in Southern Min

Ying Chen, Vsevolod Kapatsinski, Susan Guion-Anderson

Department of Linguistics, University of Oregon, USA
ychen12@uoregon.edu, vkapatsi@uoregon.edu, guion@uoregon.edu

Abstract

This paper investigates the effect of vowel quality on the perception of coda nasals in Southern Min. The perceptual confusion experiment revealed that /m/ is the most confusable coda nasal, followed by /n/ and then /ŋ/. The high front vowel /i/ resulted in more misidentification of following coda nasals than mid vowel /a/ and low vowel /a/. Within the same vowel context, higher formant frequency at the juncture of vowels and nasals and greater formant change from the midpoint to the endpoint of vowels provided more salient acoustic cues to place of articulation of post-vocalic nasals and thus resulted in higher accuracy of coda nasal identification.

Index Terms: coda nasal, perception, vowel, acoustic cues

1. Introduction

Phonological contrasts have been claimed to be distinguished based on the perception of phonetic contrasts [6]. Listeners may mishear phonetic contrasts due to ambiguous acoustic signal, possibly resulting in sound change [9]. However, the acoustic cues to the perception of phonetic contrasts exist not only in the contrasting phones themselves, but also have correlates in neighboring sounds, which are coarticulated with the target sounds [10]. Vowels have been found to bear cues to the place [2] and manner [3] of articulation of adjacent consonants. In particular, coarticulatory cues on neighboring vowels have been found to strongly influence on the perception of place of articulation for nasals [5]. The present project extends this line of investigation into Southern Min. This language is of particular interest to us, because, unlike Mandarin, it preserves /m/ from old Chinese, thus providing ideal ground to investigate possible perceptual reasons for coda nasal merger in Chinese languages. However, in this paper we focus on the issue of how perceptibility of nasal place of articulation varies across nasal consonants and vowel contexts rather than drawing connections to language change.

First, we explore the perceptual confusability of the three types of coda nasals. Bilabial nasal /m/ has been found to be more auditorily confusable in the syllable-final position than alveolar nasal /n/ and velar nasal /ŋ/, and /ŋ/ is slightly more confusable than /n/ in English [11] and French [4]. Does this perceptual confusability hierarchy /m/>/ŋ/>/n/ exist in Southern Min?

Second, we examine the effects of vowel quality on the accuracy of coda nasal identification. The high front vowel /i/ has been found to result in more, and the low vowel /a/ to result in fewer, misidentifications of coda nasals in English [11] and Mandarin [7]. Is the vowel effect on the perception of coda nasals similar in Southern Min?

Third, we investigate the acoustic cues within a vowel to the perception of nasal place of articulation. The high front vowels provide nasals with an environment in which bilabial nasals are heard as if they have a further back place of articulation [5]. Preceding the low vowel /a/, /m/-/n/ and /m/-/ŋ/ are more distinguishable than /n/-/ŋ/ due to the acoustic salience of cues to nasal place of articulation [8]. Does the acoustics of the same vowel differ before different nasals in Southern Min? Are the acoustic similarities and/or differences correlated to the identification accuracy of nasal place of articulation?

2. Methods

2.1. Stimuli

Eight real monosyllabic words in Quanzhou Southern Min were selected. They consisted of three nasal types and three vowel types, all with mid-level tone: /ɕəm33/ (心, ‘heart’), /ɕən33/ (新, ‘new’), /ɕəŋ33/ (升, ‘rise’), /ɕəm33/ (森, ‘forest’), /ɕəŋ33/ (杉, ‘student’), /ɕəm33/ (杉, ‘‘fr’), /ɕən33/ (删, ‘delete’), and /ɕən33/ (松, ‘loose’). /ɕəm33/ was absent since /m/ is not found in Southern Min sound system.

2.2. Participants

Both talkers and listeners were native speakers of Quanzhou Southern Min. There were four talkers: two males and two females. The fifteen listeners were all college students. They were familiar with the Latin alphabetic notations for /m/, /n/, and /ŋ/ as ‘m’, ‘n’ and ‘ng’.

2.3. Procedure

The eight stimulus words written in Chinese characters were presented to the talkers with five repetitions in a random order. Recording was conducted in a quiet room, using a Marantz professional solid state recorder PMD660 with a sampling rate of 44,100 Hz and a Shure professional unidirectional head-worn dynamic microphone. The stimulus words were recorded in isolation with no carrier phrases in order to avoid tone sandhi and tonal coarticulation.

The perceptual confusion experiment was conducted in a quiet room too, using a Lenovo Thinkpad PC and a Sony professional dynamic stereo headphone. The 160 target tokens (8 words * 5 repetitions * 4 talkers) were presented in a random order for a forced choice identification task by Praat. The stimulus words, whose intensity was RMS normalized, were embedded in pink noise with a +4dB SNR and played one at a time. Listeners were required to select ‘m’, ‘n’ or ‘ng’ for their identification of the word-final nasals.

Before the perception experiment, the syllable-tone sequences, i.e. the stimulus words, were presented aurally and the listeners were asked to provide a subjective familiarity rating using a 1-5 point scale. After the perception experiment, lexical familiarity ratings (1-5 point scale) were collected by visually presenting all the homophonic Chinese characters associated with each syllable-tone sequence.
3. Results

3.1. Perceptual confusability of nasals

Overall /m/ was misheard more than /n/ and /ŋ/. /n/ was most accurately identified among the three coda nasals. There were more /m/-/ŋ/ confusions than /m/-/n/ and /n/-/ŋ/ confusions. Table 1 shows the results of the perceptual confusion experiment for the three nasals preceded by the three vowels /i, ə, a/.

Table. Confusion matrix of coda nasals by vowel types (n = 20 tokens * 15 listeners = 300).

<table>
<thead>
<tr>
<th>Spoken</th>
<th>Heard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m</td>
</tr>
<tr>
<td>eam</td>
<td>241</td>
</tr>
<tr>
<td>en</td>
<td>28</td>
</tr>
<tr>
<td>eŋ</td>
<td>51</td>
</tr>
<tr>
<td>em</td>
<td>167</td>
</tr>
<tr>
<td>eŋ</td>
<td>36</td>
</tr>
<tr>
<td>em</td>
<td>129</td>
</tr>
<tr>
<td>en</td>
<td>40</td>
</tr>
<tr>
<td>eŋ</td>
<td>59</td>
</tr>
</tbody>
</table>

Because of the absence of /æŋ/, two repeated measures ANOVAs were conducted to test the effect of vowel type on the identification of coda nasals. The first ANOVA with factors of nasal /m, n, ŋ/ and vowels /i, ə, a/ showed a significant interaction of nasal and vowel [F(2,28) = 9.153, p = 0.001]. The second ANOVA with factors of nasal /m, ŋ/ and vowels /i, ə, a/ again showed significant interaction of nasal and vowel [F(2,28) = 6.620, p = 0.004]. The source of the interactions was that /m/ was less correctly identified than /n/ and /ŋ/ in /i/ and /ə/ contexts, whereas /ŋ/ was least well identified in the /a/ context. Nasals following /i/ were less correctly identified than following other vowel types. Post hoc paired-samples T tests show that /m/ was less correctly identified than /n/ in both /a/ context [ t(1,14) = -4.089, p = 0.001] and /i/ context [ t(1,14) = -4.319, p = 0.001], and than /ŋ/ in /i/ context [ t(1,14) = -3.860, p = 0.002]; /i/ resulted in less correct identification of /m/ [ t(1,14) = -4.319, p = 0.001] and /n/ [ t(1,14) = -4.089, p = 0.001] than /ŋ/ did.

Pearson correlations showed no significant relationship between the identification accuracy of word-final nasals and audio familiarity ratings of phonetic wordforms, written familiarity ratings of orthographic wordforms or the number of homophones having the same wordform.

Figure 1: Mean frequency of vowel formants at midpoint, endpoint and formant change for the five repetitions by the four talkers.
3.2. Acoustic analysis of vowels

The results of nasal identification task show an effect of vowel contexts on the perception of coda nasals. Due to the lack of /ɕən/, the acoustic analysis only involves vowels /a, i/ by nasals /m, n, ŋ/. Vowel duration and nasal murmur duration were measured. However, there were no significant correlations between the accuracy of nasal identification and either vowel duration, nasal duration or the sum duration of vowel and nasal.

Figure 1 shows the acoustic characteristics of the formant transitions from the midpoint to the endpoint of vowels by nasal type. When boxes do not overlap, the difference is significant. No difference was found between nasals in F1 at vowel midpoint. F2 of /a/ before /n/ rose more than before other nasals: /a/ was a bit fronted before /ns/ even at the midpoint and very front by the end right before /ns/. /a/ and /i/ differed on F2 change of either /a/ or /i/; F2 change before /ŋ/ was intermediate compared to F2 change before /n/ and F2 change before /n/ but was closer to F2 change before /ns/. F3 endpoint of /i/ was lower before /m/ than before /n/. /a/ and /i/ seemed to differ on F3 change, especially after /i/; /i/ was like /n/ on F3 change.

3.3. Correlation between vowel quality and nasal perception

Linear correlation between the accuracy of nasal identification and the values of midpoint, endpoint and change of the three vowel formants was investigated. Because distributions of accuracy scores were skewed, accuracy scores were converted to log scale [log (change – minimum change +1)] instead of raw scores. Table 2 shows the correlation results.

Table 2. Correlation between nasal identification accuracy and vowel formant frequency (df = 358; significance at a Bonferroni-adjusted level of 0.0028).

<table>
<thead>
<tr>
<th></th>
<th>Midpoint</th>
<th>Endpoint</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>t = 5.984</td>
<td>t = 6.274</td>
<td>t = 4.186</td>
</tr>
<tr>
<td></td>
<td>p &lt; 0.001*</td>
<td>p &lt; 0.001*</td>
<td>p &lt; 0.001*</td>
</tr>
<tr>
<td>F2</td>
<td>t = -2.13</td>
<td>t = 2.109</td>
<td>t = 6.583</td>
</tr>
<tr>
<td></td>
<td>p = 0.007</td>
<td>p = 0.036</td>
<td>p &lt; 0.001*</td>
</tr>
<tr>
<td>F3</td>
<td>t = -2.33</td>
<td>t = 6.024</td>
<td>t = 7.700</td>
</tr>
<tr>
<td></td>
<td>p = 0.020</td>
<td>p &lt; 0.001*</td>
<td>p &lt; 0.001*</td>
</tr>
</tbody>
</table>

Table 2 indicates that the accuracy of nasal identification is significantly correlated to F1 midpoint, F1 endpoint, F1 change, F2 change, F3 endpoint, and F3 change. Figure 2 plots these correlations. All significant correlations are positive.

Figure 2: Scatter plots of the numbers of correct nasal identification as a function of vowel formants and their change.
4. Discussion

The perceptual confusability hierarchy /m/ > /ŋ/ > /n/ was found in Southern Min just as in English [11] and French [4] with the difference between /ŋ/ and /n/ being not as large as that between /m/ and /ŋ/. From this result alone we cannot yet say whether the observed differences are due to differences in acoustic cue availability and salience or to phonotactics of Southern Min. However, consider that /m/ is a phonotactically illegal rime in Southern Min. Nonetheless, our listeners often reported hearing /mən/ when presented with /məm/ or /məŋ/. This finding suggests that they identified coda nasals based largely on the acoustic signal and were not strongly affected by phonotactic biases. That the listeners were relatively unaffected by lexical biases was also reflected in the non-significance of correlations of identification accuracy with either subjective lexical familiarity or word type frequency. These results suggest that we should look to acoustics for explanations of the perceptual confusability hierarchy and vowel context effects in the present study.

Since vowel duration was not correlated to the accuracy of nasal identification, vowel formant structures can be considered as the primary acoustic cues to the perception of coda nasals. F1 midpoint marks vowel height and thus represents vowel type. We can see from the acoustics (Figure 1) that F1 midpoint varied greatly across vowels but not across nasals within the same vowel. F1 midpoint is unlikely to provide acoustic cues to nasal place of articulation but correlates to the overall accuracy of nasal perception. This can be seen in the result that the nasal identification was more accurate after /a/ than after /i/ in the current study, as in the literature [7, 11]. More /m/-/n/ confusion than /m/-/a/ and /n/-/n/ confusion could be caused by the similar formant structure, especially F1 change and F2 change, in /a/ and /i/ contexts. High F2 endpoint and great F2 change resulted in a fronted /a/ before /n/ and thus a high accuracy of /n/ perception in /an/. Compared to the diphthongization of /i/ before /m/ and /n/, no noticeable frequency change of all the three formants for /i/ before /n/ brought a higher accuracy of /n/ identification than that of the other two nasals. Compared to other vowel contexts, /i/ greatly decreased the accuracy of /m/ identification. /i/ did not have such a strong effect on identification accuracy of /n/ or /ŋ/. This result is consistent with the conclusion in [5] that high front vowels make /m/ likely to be misperceived further back. The major acoustic effect of /i/ context on cues to /m/ in the present study, given than the stimuli’s intensity was RMS-normalized, was the great reduction in F3 at the vowel-nasal boundary. However, it is unclear why this should lead to misperceiving /m/ as being articulated further back. Finally, the perception of /ŋ/ was relatively unaffected by preceding vowel. The main cue to the velar place of articulation is the velar pinch, which was similar in magnitude across vowels.

5. Conclusions

The results of the current study indicated that coda /m/ is perceptually more confusable than coda /n/ and /ŋ/ in Southern Min. The high front vowel /i/ resulted in more misidentification of coda nasals than mid vowel /a/ and low vowel /u/. Higher formant frequency at the juncture of vowels and nasals and greater formant change from the midpoint to the end point of vowels resulted in higher accuracy of nasal identification in the same vowel context.

This study only involves preceding vowel formant structures as acoustic cues to place of articulation of coda nasals. However, the perceptual cues to nasal place of articulation could also be related to other acoustic features, such as the amplitude of vowel formants and nasal peaks [1, 7]. This issue is to be addressed in future work.

6. References