NASAL CODA REALISATION IN SPEECH PRODUCTION OF SHANGHAI MANDARIN

Yihan GUAN

City University of Hong Kong
masaharuguan@gmail.com

ABSTRACT

This study investigated the nasal coda realisation in Shanghai Mandarin where speakers sometimes confuse alveolar and velar nasal codas to see whether it would be influenced by the phonetic environment and speech styles. Twenty-five participants from Shanghai were involved in two speech production tasks. The analyses showed that the nasal neutralisation mainly happened from velar nasals to alveolar counterparts, and it exhibited more with the presence of the preceding vowel /i/, then /ə/, and /a/ was the least possible to trigger such phenomenon. In addition, the confusion would favour casual speech more than formal speech, but each style exhibited different neutralising patterns.

Keywords: nasal coda; Shanghai Mandarin; speech production; language variation and change

1. INTRODUCTION

In a Chinese (Mandarin) syllable, the coda position can be filled with a nasal contrast, /n/ and /ŋ/, and they could be both preceded by three monophthong vowels, /i/, /ə/ and /a/. For /a/, there are two allophones, the back vowel [ɑ] only appears when preceding a velar nasal, and [a] can appear with the presence of an alveolar nasal coda. For /ə/ and /i/, in Shanghai Mandarin, realisation before each nasal is the same.

Given the variants available from the preceding vowels, we can see that in pair [an] - [ɑŋ], there are at least two cues available to distinguish each other, namely the preceding vowel and the nasal, but in [ən] - [əŋ] as well as in [in] - [iŋ], the only obvious cue is the nasal difference. According to the phonologization theory [8], if there exist two cues to distinguish two contrast sounds from each other, the robust one would serve as the primary cue and the other as the secondary (redundant) cue, and the informative primary cue is more likely to be selected while the redundant cue may be reduced due to its obscureness [21]. As nasal is a relatively weak cue [9] compared to the vowel difference, theoretically the neutralisation should happen more in nasals preceded by /a/, resulting in a pair of [an] - *[aŋ], or *[ɑn] - [ɑŋ], but for /ə/ and /i/, nasal difference is the only element that distinguishes the syllable, and therefore however weak it is, the nasal difference should be retained.

However, recent studies on the nasal coda neutralisation do not suggest so. There have been little done in Shanghai Mandarin but those mainly focusing on Taiwan Mandarin show that the neutralisation goes from velar nasals to the alveolar counterparts, and the neutralisation rate is the lowest preceded by /a/, increases when followed by /ə/, and becomes the easiest to change when preceded by the high front vowel /i/ [6, 10, 14, 15, 21].

Different speech styles could also lead to a difference in the actual realisation of syllable-final nasals. In studies in Taiwan Mandarin, subjects tend to confuse alveolar and velar nasals more in spontaneous (casual) speech. In [7] which focused on Taiwan Mandarin, spontaneous speech would result in more lack of nasal contrast. However, in [15], there seems to be no significant difference in regard to speech styles.

Given the contradictory results found in different studies as well as a lack of such studies on Shanghai Mandarin, this study aims at investigating the following questions:

1. Do speakers from Shanghai produce /n/ and /ŋ/ differently with different preceding vowels?
2. Do different speech styles show different patterns?

2. METHODS

In order to investigate whether realisation would vary with different preceding vowels and styles, two experiments were designed. All participants were involved in two tasks. After filling a demographic questionnaire, they would do an interview first, and then a wordlist reading task.

2.1. Participants

A total of 25 participants were recruited for the current study. They were all students at the University of Edinburgh, recruited through the "friend of a friend" approach [16]. Among all the participants, 7 were male and 18 were female, aged between 22 and

28 (M = 23.84, SD = 1.76). They were raised in Shanghai (born in Shanghai or lived in Shanghai before 4 years old), and the average duration of residence in Shanghai was 21.13 years (SD = 2.93), meaning that they have spent on average 89% of their lives living in Shanghai. Twenty of them were self-reported to be able to speak Shanghainese, a regional dialect in the area (either as their first language or one of their first languages). All of them reported no speech impairments.

2.2. Equipment

All participants were recorded in a quiet room on the premises of the University of Edinburgh using Zoom H2n recorder, digitised at 44.1khz and 16-bit.

2.3. Materials

In the first production task aiming at eliciting casual speech in a face-to-face interview, twelve semantic differential questions which asked about differences between a pair of synonyms were arranged. Four of them were distractors and eight were of interest. They were all displayed on the screen pair by pair. Each pair consisted of real and high-frequency words, and all target syllables were in CVN structure containing either /a/, /ə/ or /i/ and one of the nasals in the coda. Target tokens could be in any position of a word given that spontaneous speech did not have a clear pause between syllables. For example, in one of the semantic differential questions, the two synonyms were: ***bing*** *qi* ***lin*** "ice cream" and ***bing bang*** "ice lollies", with bold syllables targeted. Nineteen syllables in total were intended.

In the second production task, which used a wordlist to elicit formal speech produced with more attention, a total of 180 real words were arranged, including 72 fillers and 108 target words. All 108 target words were real and high-frequency words, and target syllables were in CVN structure placed in the word final position which was followed by an interval to avoid influence from the next syllable. For example, token *huang* ***jin*** "gold" was used in this task, with the bold syllable targeted. Eighteen syllables were evenly distributed for each vowel and nasal type (18 * 3 vowels * 2 nasals = 108 tokens).

2.4 Procedure

After signing the consent form and completing the demographic questionnaire, participants would sit in front of a computer. For the first task, each pair of semantical synonyms were displayed in the centre of the screen in a randomised order using PsychoPy [18], and they were displayed instead of being said by the interviewer to avoid the influence from the interviewer. Participants would discuss the differences in meaning between them with the interviewer, and the discussion for each pair usually lasted for 2-3 minutes. After the first task and a one-minute break, participants would come to the wordlist reading task. A practice session with five trials was given for participants to familiarise with this task before the official start. Stimuli would appear in the centre of the screen in a randomised order using PsychoPy [18], and participants had to read out the word they saw. An inter-trial-interval of 3,500 ms was placed. A 30-second break was provided in the middle of this task.

2.5. Data labelling

All tokens were manually judged and labelled by the researcher. Acoustic analysis was carried out with the aid of Praat [2] in combination with the perceptual judgment of the researcher who could distinguish two nasals in both production and perception. F2 at the endpoint of preceding vowels was accepted as a strong cue in judging how the nasal coda was realised [12] as during the transition of the front vowel [i] to the velar nasal, F2 would move up and F3 would move down to merge with F2 to form a velar pinch, but for the transition to the alveolar nasal, the trajectories of F2 and F3 would keep parallel. However, for mid [ə] and [a], and back vowel [ɑ], the transition to velar nasals would not display an obvious velar pinch [17], but just a slight rise of F2 at the end, therefore, auditory judgment was the primary source of nasal identification in combination with F2 values at the end of preceding vowels.

3. RESULTS

An inspection of all tokens extracted was carried out before analysis. From the first task, a total of 2,297 tokens were extracted, of which 2,152 tokens were included for analysis. The exclusion of 145 tokens was to prevent the influence of the first token on tokens consecutively repeated in a row, therefore tokens in a repetition string other than the first one would be excluded. In addition, tokens whose vowels and/or nasal codas were too short in duration to be identified in connected speech would also be excluded. From the second task, all tokens (108 * 25 = 2700) from each participant were clearly and correctly pronounced and recorded.

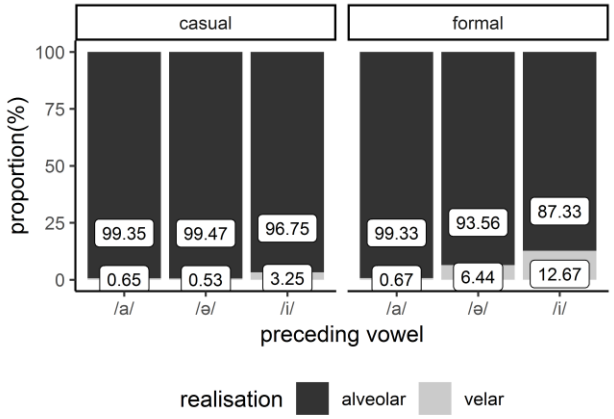
A mixed-effects logistic regression was originally built in R [19] with the *lme4* package [1] in order to investigate the relationship between the actual realisation and intended nasals when modulated by preceding vowels as well as styles. However, the model failed to converge when all data were entered,

therefore two models based on different intended nasal types were built, and both models converged this time. The dependent variable for each model was the agreement between the actual realisation and the intended nasal (yes, no). As fixed effects, preceding vowel and style were entered with the preceding vowel /i/ and the casual style as the intercepts. Gender and whether participants could speak Shanghainese were also entered as control variables. As previous studies on nasal realisation in Taiwan Mandarin indicated that the lack of nasal contrast in Min dialect [10] lead to the confusion of nasal codas in Taiwan Mandarin, therefore, Shanghainese, the regional dialect where only the velar nasal [ŋ] is allowed in the coda position [4] was controlled in the current study. The interaction between vowel and style was treated as the interaction item. As random effects, participants and different stimuli were entered as intercepts. P values were obtained using the *lmerTest* package [11]. Pairwise comparison using Tukey adjustment was done using emmeans function in the *emmeans* package [13] when significant effects were shown.

3.1 Intended alveolar nasals

Figure 1 shows that velarisation of intended alveolar nasals varies with different preceding vowels.

Figure 1: Actual realisation of intended alveolar nasals following different vowels



When preceded by /a/, both styles present merely 0.65% and 0.67% mismatch between the intended alveolar nasals and the actual realisation, and 0.53% for /ə/ in casual speech, but higher in terms of formal speech, 6.44%. For the preceding vowel /i/, around 3.25% of alveolar nasals are realised as velar nasals in casual speech and a higher 12.67% in formal speech. The neutralisation rate mostly follows the order /i/ > /ə/ > /a/ (high to low).

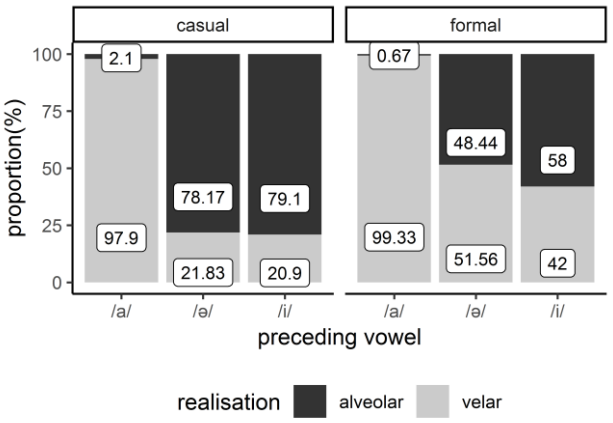
The logistic mixed-effects analysis shows that different preceding vowels influence the realisation of alveolar nasals. Realisation with the presence of preceding vowel /i/ is significantly different from that of /a/ or /ə/ (b = 1.68 and 1.83, SE = 0.65 and 0.76, p < 0.01 and 0.05, respectively), but a pairwise comparison between preceding /a/ - /ə/ indicates no significant difference (b = 1.13, SE = 0.54, p = 0.09). Style difference is statistically significant (b = -1.6, SE = 0.34, p < 0.001), but gender or ability to speak Shanghainese does not influence the realisation of alveolar nasals (b = 0.53 and 0.03, SE = 0.88 and 0.86, p = 0.5138 and 0.9681, respectively).

Multiple pairwise comparisons of preceding vowels modulated by styles show that the results exhibit more difference within formal speech as all three preceding vowel comparison pairs show statistical difference (all $|z|$ > 3.3, p < 0.003) as the rate increases from /a/ to /ə/ then to /i/, while the difference is less significant in casual speech, with pair /i/ - /a/ and /i/ - /ə/ showing marginal significant (both $|z|$ > 2.3, p < 0.04) but not for pair /a/ - /ə/ ($|z|$ = 0.15, p = 0.98) as the rates are both around 0.5%.

3.2 Intended velar nasals

Figure 2 shows that different preceding vowels lead to different patterns in realising intended velar nasals.

Figure 2: Actual realisation of intended velar nasals following different vowels



With preceding /a/, 0.67% of velars are realised as the alveolar counterparts in formal speech, 2.1% in casual speech. For /ə/ the mismatch rates increase drastically to 48.44% in formal speech and 78.17% in casual speech. A similar trend is observed with preceding /i/ where the neutralisation rates are 58% and 79.1% respectively in formal and casual speech. The neutralisation rate therefore follows the order /i/ > /ə/ > /a/ (high to low).

The logistic mixed-effects analysis indicates that the neutralisation rate of alveolar nasals to the velar counterparts would differ with different preceding vowels. The difference between nasal realisation following /i/ - /a/ is significantly large (b = 7.52, SE = 1.02, p < 0.001), so as between /a/ - /ə/ (b = 6.39,

SE = 0.60, $p < 0.001$), but no difference is shown in /i/ - /ə/ pair (b = 0.05, *SE* = 0.20, p = 0.81). Style difference is statistically significant in terms of intended velar nasals (b = 1.17, *SE* = 0.17, $p < 0.001$), but gender or ability to speak Shanghainese does not influence the realisation of nasal codas significantly (b = 0.90 and 0.72, *SE* = 0.66 and 0.62, p = 0.17 and 0.24, respectively).

Multiple pairwise comparisons of velar nasals preceded by different vowels modulated by styles show less difference within formal speech, contradicting to intended alveolar nasals case. In formal speech, alveolarisation occurs differently when preceded by different vowels in all three pairs (all $|z| > 3.1$, $p < 0.005$), as the neutralisation rates increase from /a/ to /ə/ and gradually to /i/, but in casual speech, two pairs that show significant difference are /i/ - /a/ and /a/ - /ə/ (both $|z| > 7.2$, $p < 0.001$), but not /i/ - /ə/ ($|z| = 0.24$, p = 0.97), as the neutralisation rates of this pair are in two styles similarly high around 79%.

4. CONCLUSION AND DISCUSSION

This study found that the nasal coda neutralisation in Shanghai Mandarin is mostly from velar nasals to alveolars, but the opposite direction also occurs, which depends on the speech style. Both directions follow the same pattern: /i/ > /ə/ > /a/.

The reason for the phenomenon to appear could to some extent be attributed to the internal structure of nasal syllables. Speakers do not confuse two nasals in the production of the pair [an] - [ɑŋ] because of the existence of the stronger cue [a] - [ɑ] contrast, rather, both vowel and nasal cues in conjunction enhance the [an] - [ɑŋ] distinction, while for [ən] - [əŋ] and [in] - [iŋ] pairs, although two syllables would no longer be distinguishable without nasal difference, speakers nonetheless continue the neutralisation. This would more possibly be the result of following Rhyme Harmony [5], which constraints that the nucleus and coda must agree in backness. For [an] and [ɑŋ], the place of articulation already agrees ([-back][-back] and [+back][+back], respectively), and the violation cases *[aŋ] and *[ɑn] are not preferred in Mandarin phonology. For /ə/, as the non-high vowel position is unmarked for backness, the combination with nasals would be unconstrained, hence sometimes the nasal neutralisation is favourably disposed to but sometimes the distinction wins out. For preceding /i/, [in] agrees [-back][-back] but [iŋ] is a violation in terms of the Harmony, therefore, lower accuracy is observed in it in both styles.

The physical account for Rhyme Harmony is the capability for an articulator to move fast enough to execute two opposite gestures. Among all combinations of three vowels and two nasals, the longest distance for the articulator to move is [iŋ], which is from the very front of the oral cavity to the velum, therefore, the difficulty, as well as the excessive energy spent in realising it, would intervene in the distinction the most. But it cannot explain why the nasal neutralisation from velar to alveolar happens in Shanghai, but not in Beijing, where such neutralisation goes the other way around when preceded by /i/ [3]. Therefore, future studies could possibly consider factors inhibiting such neutralisation in Beijing Mandarin if Rhyme Harmony should play a role in the neutralisation.

The stylistic variation should be interpreted with caution. For the velarization of the intended alveolar nasals, however unfavourable the pattern is (mismatch rate generally is below 15%), we can see that such phenomenon persists more in formal speech, especially with preceding vowel /i/ as the mismatch rate increases by almost 10 percent points from 3.3% in casual speech to 12.7% in formal speech. For the alveorisation of the intended velar nasals, although the rate remains low for preceding vowel /a/ in both styles, when preceded by /ə/ and /i/, the neutralisation rate decreases from 78.2% and 79.1% in casual speech to 48.4% and 58% in formal speech.

Two possibilities are proposed for the above stylistic variation. The first interpretation is that speakers possess phonological knowledge of two nasal codas, and the problem lies only in phonetic realisation. Therefore, when more time and attention is allowed in formal speech, chances are high that they could successfully realise intended velar nasals, therefore we see an overall lower mismatch rate with intended velar nasal codas in formal speech than in casual speech. A second explanation could be that speakers are aware of the existence of two nasal codas through formal education, media and other settings, and they are also reminded of their excessive use of alveolar nasals in Shanghai Mandarin, which is considered prescriptively wrong, but they do not phonologically distinguish the two phonemes. Therefore, when more time is given, they just hypercorrectly velarise possible alveolars, which results in the unusual higher mismatch rate with intended alveolar nasals in formal speech as compared to those realised in casual speech.

Future studies could include language attitudes as an external factor behind the phenomenon. It is possible that they are aware of the prescriptive distinction, but they nonetheless prefer to speak the variety without nasal coda differentiation that is different from varieties spoken in Beijing and other northern areas in China.

5. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

[2] Boersma, P., Weenink, D. 2018. Praat: Doing Phonetics by Computer [Computer program]. Version 6.0.40, http://www.praat.org/.

[3] Chen, C. Y. 1991. The Nasal Endings and Retroflexed Initials in Peking Mandarin: Instability and the Trend of Changes. *Journal of Chinese Linguistics, 19*(2), 139-71.

[4] Chen, Y., Gussenhoven, C. 2015. Shanghai Chinese. *Journal of the International Phonetic Association, 45*(3), 321-337.

[5] Duanmu, S. 2007. *The Phonology of Standard Chinese.* Oxford: Oxford University Press.

[6] Hsu, H-J., Tse, J. K. 2007. Syllable-Final Nasal Mergers in Taiwan Mandarin: Leveled but Puzzling. *Concentric: Studies in Linguistics, 33*(1), 1-18.

[7] Hung, C. W. 2006. A Sociolinguistic Variation Study of Taiwan Mandarin Velar Nasal in Kaohsiung. MA Thesis, Tainan: National Cheng Kung University. http://etds.lib.ncku.edu.tw/etdservice/view_metadata ?etdun=U0026-0812200911502261.

[8] Hyman, L. M. 1976. Phonologization. In A. Juilland, A. M. Devine, & L. D. Stephens (Eds.), *Linguistic studies offered to Joseph Greenberg on the Occasion of His Sixtieth Birthday*, Vol. 4. Anima Libri, 407-18.

[9] Kawahara, S., Garvey, K. 2014. Nasal Place Assimilation and the Perceptibility of Place Contrasts. *Open Linguistics, 1*, 17-36.

[10] Kubler, C. C. 1985. The Influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics, 27*(2), 156-176.

[11] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26. doi: 10.18637/jss.v082.i13.

[12] Ladefoged, P. 2003. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques.* Oxford: Blackwell Publishing Ltd.

[13] Lenth, R. 2018. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.2.3. https://CRAN.R-project.org/package=emmeans.

[14] Li, A., Fang, Q., Xu, R., Tang, Y. 2005. A Contrastive Study Between Minnan-Accented Chinese and Standard Chinese. *Report of Phonetic Research, 18*, 1-7.

[15] Lin, C. C. 2002. Nasal Endings of Taiwan Mandarin: Production, Perception, and Linguistic Change. Paper presented at the 35th International Conference on Sino-Tibetan Languages and Linguistics. Tempe: Arizona State University.

[16] Milroy, L. 1980. *Language and Social Networks*. London: Basil Blackwell.

[17] Olive, J. P., Greenwood, A., Coleman, J. 1993. *Acoustics of American English Speech: A Dynamic Approach.* New York: Springer.

[18] Peirce, J. W. 2007. PsychoPy - Psychophysics Software in Python. *Neurosci Methods, 162*(1-2), 8-13.

[19] R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[20] Steriade, D. 1997. Phonetics in phonology: the case of laryngeal neutralization. *UCLA Working Papers in Linguistics, 3*,125-46.

[21] Tse, J. K. 1992. Production and Perception of Syllable Final [n] and [ŋ] in Mandarin Chinese: An Experimental Study. *Studies in English Literature and Linguistics, 18*, 143-56.