PRODUCTION AND PERCEPTION OF MANDARIN /I/-NASAL RHYMES

Suyuan Liu and Molly Babel

University of British Columbia
suyuan.liu@ubc.ca, molly.babel@ubc.ca

ABSTRACT

This study compares how speakers of Beijing Mandarin (BjM) and Shanghai Mandarin (ShM) differ in their production and perception of nasal codas as a means of evaluating the (mis)alignment of perception and production. Speaker-listeners completed a word-level production task, a word recognition task, and a goodness rating task. Results suggest that BjM users contrast /in/-/iŋ/ words in perception and production by consistently mapping [in] and [iɲ] to the former and [iŋ] and [iəŋ] to the latter. ShM users show no /in/-/iŋ/ word contrast in production while consistently mapping the [iəŋ] variant to /iŋ/-words and rating it as the best /iŋ/-exemplar in perception. This indicates a mismatch in perceptual and production, providing insights into cross-dialect variation.

Keywords: nasals, production, perception, Mandarin

1. INTRODUCTION

Mandarin includes diverse varieties that emerge through the promotion of standard Mandarin to speakers of Chinese languages. Different Mandarin varieties have distinct production variants that may be unfamiliar to speakers of other varieties. However, the term "Mandarin" obscures multiple language concepts [1]. Two terms are used in this paper to differentiate the concepts. *Standard Mandarin* is used to refer to the idealized Mandarin promoted in China. *[City Name] Mandarin* (e.g. Shanghai Mandarin) is used in reference to the Mandarin variety spoken in or around the city region, which is *not* the same as the local Chinese language (e.g. Shanghainese).

Standard Mandarin only allows two codas /n/ and /ŋ/. The degree of distinction between /n/ and /ŋ/ is conditioned by three main factors: (1) the preceding vowel, (2) the Mandarin variety, and (3) whether the investigation is in speech perception or production [2, 3, 4, 5]. Among all Mandarin varieties, Taiwan Mandarin (TwM), Shanghai Mandarin (ShM) and Beijing Mandarin (BjM) are the most well-studied when it comes to Mandarin nasal codas. In all three varieties, the degree of distinction in production and perception is vowel-dependent. The high front vowel rhyme /iN/ – N is used to represent an unspecified nasal – has the lowest degree of distinction in TwM and ShM production [3, 5, 6, 7, 8, 9, 10] and the lowest perceptual classification accuracy [3, 4, 5, 11]. In production, both groups use a (fronted) velar tongue gesture when producing the nasal in /iN/ [3, 6]. BjM, however, makes such a distinction by realizing /in/ as [in] and /iŋ/ as [iəŋ], a foreign variant to ShM and TwM [12, 13, 14]. But interestingly, even BjM listeners are poor at perceiving /iN/ rhymes [3, 15, 16], which suggests a misalignment between perception and production. However, the perception studies did not take into account the vowel differences that are present in BjM production. In addition, the majority of previous perception studies focused on phonetic distinction [3, 16] or instructed listeners to map phonemes to a Pinyin representation (e.g. select "n" or "ng" for the nasal you heard) [5, 15]. It is unclear whether the poor perception performance generalizes to a lack of sensitivity at the lexical level. To address the perception and production misalignment in [3], the current work reassesses the nasal coda distinction in perception and production of /iN/ rhymes when considering the vowel differences and focuses perceptual sensitivity to the word level for BjM and ShM.

2. GENERAL PROCEDURE

Participants were recruited through Chinese social media from Beijing and Shanghai. 29 participants' data were analyzed: 14 in the ShM group (9 male, 5 female, M = 31.9 years, range = 22-56) and 15 in the BjM group (7 male, 8 female, M = 26.7 years, range = 20-32). All participants completed language background and residential history questionnaires and a headphone [17] and microphone check. They then took part in a production experiment and a set of perception experiments: a two-way alternative forced choice task and a goodness rating task with Chinese instructions. Including both tasks examines category boundaries and corresponding prototypes, which could differ by dialectal groups [18]. All experiments were hosted online using Gorilla [19].

3. PRODUCTION STUDY

Participants were instructed to record themselves producing the Chinese character shown on the screen. Four practice trials were completed prior to the actual task. Trials were separated by a 450ms fixation pause. The screen automatically advanced to the next trial if no recording was made in 3s.

3.1. Material

Production stimuli consisted of 9 minimal pairs of frequency-matched monosyllabic words in Chinese characters. Each character appeared twice per block for three blocks (totaling 108 trials). The 3 pairs of target stimuli are shown on the left in Table 1. The other 6 pairs were codaless fillers items (e.g. [pʰwo] ‘splash’ 泼- [pwo] ‘spread’ 播).

Production Stimuli	Perception Stimuli			
	in	iɲ	iŋ	iəŋ
/jin/-/jiŋ/ ‘sound’ - ‘eagle’ 音- 鹰	[jin]	[jiɲ]	[jiŋ]	[jiəŋ]
/bin/-/biŋ/ ‘guest’ - ‘ice’ 宾- 冰	[bin]	[biɲ]	[biŋ]	[biəŋ]
/tɕin/-/tɕiŋ/ ‘gold’ - ‘shock’ 金- 惊	[tɕin]	[tɕiɲ]	[tɕiŋ]	[tɕiəŋ]

Table 1: Target production and perception stimuli

3.2. Production coding

All tokens were manually coded by a Spanish-English bilingual and a Northern Mandarin-English bilingual. Both coders are phonetically trained. Coders listened to all randomized tokens and labeled each token as (1) [in], (2) [iŋ], (3) [iəŋ], (4) [iɲ], (5) no nasal coda, or (6) other with text for specification. Two coders agreed on 622 out of 1149 tokens (unweighted Cohen’s kappa = 0.581, $p < 0.001$) and jointly relabeled all tokens disagreed on.

3.3. Production results

Impressionistically, the distribution of /iN/-words pronunciations show similarity between BjM and ShM groups. To assess this and facilitate model interpretation, separate models were run for the /in/ and /iŋ/ words, allowing for cross-dialect comparisons within each lexical set. Two Bayesian mixed-effect multinomial logistic regression models were fit separately for /in/ and /iŋ/ words in Stan

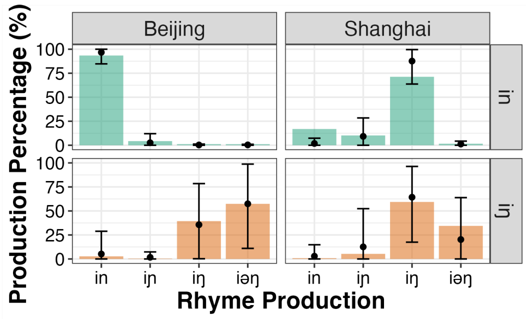

Figure 1: Participants’ proportion of produced categories when presented with a character for /in/-word or /iŋ/-word (rows) for the two dialects (columns). The error bars show the 95% HDI of posterior predictions for each group.

using the {brms} package [20] in R [21]. Both models included an outcome variable of dummy-coded Production Category ([in], [iŋ], [iɲ], and [iəŋ]; ref level: [iŋ]) and a predictor variable of dummy-coded Mandarin varieties (ShM vs. BjM; ref level: BjM). Both models also included by-speaker and by-word random intercepts. Samples were drawn from the posterior distribution using a four-chain Hamiltonian Monte-Carlo sampling (2000 iterations, 100 warm-ups). Weakly informative priors were used for all parameters. The mean of the posterior distribution, the 95% credible interval (CrI), and the probability of direction (PD) are reported. A CrI that does not encompass 0 suggests a meaningful effect and a PD greater than 95% suggests a probable direction of effect [22]. The bars in Figure 1 show the average percentage of producing one of the four variants when seeing a /in/-word or /iŋ/-word for each Mandarin variety in the empirical data. The error bars show the range of posterior predictions for each grouping.

The models give evidence that BjM speakers distinguish /iN/ by consistently producing [in] tokens for /in/-words (β = 7.46, CrI = [3.57, 11.81], PD = 99.95%) while producing a mix of [iŋ] and [iəŋ] (β = 0.61, CrI = [-1.85, 3.11], PD = 94.77%) for /iŋ/-words. For ShM speakers, the model gives evidence that they consistently produce more [iŋ] for /in/-words than [in] (β = -5.95, CrI = [-9.42, 0.08], PD = 97.32%), and produce more [iŋ] tokens than [iəŋ] tokens (β = -1.6, CrI = [-4.57, 0.11], PD = 97.08%) for /iŋ/-words.

4. PERCEPTION: LEXICAL CATEGORIZATION

In this two-way alternative forced choice task, participants heard a stimulus and saw two Chinese characters for a minimal pair on the screen. Left and right arrow keys were used to indicate which character they thought they heard. Participants started with four practice trials with feedback. No feedback was given after the practice trials. The screen advanced to the next trial if no answer was submitted in 2.5s. A 1-sec fixation reset pause was added between trials.

4.1. Materials

Perception stimuli consisted of 4 production variants of the same 3 minimal pairs listed in Table 1 and 2 productions of the canonical endpoints of 6 filler minimal pairs. All tokens were produced by a 24-year-old female Northern Mandarin-speaking trained linguist who grew up in Shandong, China. Stimuli were digitally recorded at a sampling frequency of 44.1 kHz. All production tokens were then coded by two Spanish-English bilingual trained linguists as [n], [ŋ], [ɲ], or [ɴ]. Only tokens that both coders agreed to match the target production were selected as stimuli. Two tokens were used for each variant. Each auditory stimulus was repeated twice per block for two randomized blocks (totaling 192 trials, 8 trials for each lexical target). The intensity of all stimuli was scaled to 75 dB and .wav files were converted to the Ogg vorbis container format.

4.2. Categorization task results

Within dialect group quantification was made to evaluate the perceptual equivalence of the pronunciation variants. Two Bayesian mixed-effect logistic regression models were fit separately for ShM and BjM groups using the {brms} package [20] in R [21] with the dummy-coded Probability of Velar Responses (/iŋ/ = 1, /in/ = 0) as the outcome variable. The dummy-coded auditory Stimuli Type ([in], [iŋ], [iɲ], [iəŋ]; ref level: [iŋ]) was the predictor variable. A by-subject random intercept with Stimuli Type as a by-subject slope and a by-word random intercept were included. Samples were drawn from the posterior distribution using a four-chain Hamiltonian Monte-Carlo sampling (2000 iterations, 500 warm-ups). Weakly informative priors were used for all parameters. The bars in Figure 2 show the average percentage of responses where a velar vs. alveolar character is selected in response to the auditory stimuli. The error bars show the mean and range of posterior predictions of velar responses for each stimuli type.

The models give evidence that BjM listeners distinguish /in/-/iŋ/ by consistently choosing /in/-words for [in] (β = -4.58, CrI = [-8.66, -4.76], PD = 100%) and [iɲ] (β = -0.82, CrI = [-3.59, -1.80], PD = 100%), and /iŋ/-words for [iŋ] (β = 1.82, CrI = [-0.03, 3.64], PD = 97.37%) and [iəŋ] (β = 5.08, CrI = [1.99, 5.51], PD = 100%). ShM listeners only distinguish /in/-/iŋ/ by consistently selecting /iŋ/-words for [iəŋ] (β = 2.47, CrI = [1.82, 4.05], PD = 100%) and /in/-word for the rest. The posterior distributions in Figure 2 also show that BjM listeners are more likely to choose /in/-words for [in] than [iɲ], and more likely to choose /iŋ/-words for [iəŋ] than [iŋ].

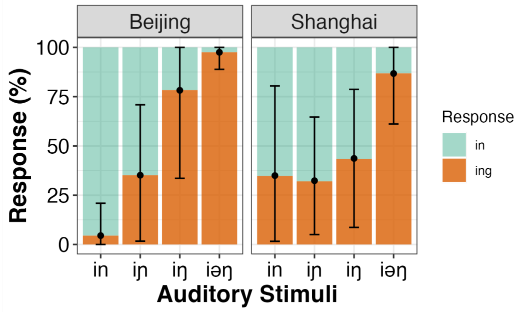

Figure 2: Word categorization results for ShM and BjM by stimuli type (x-axis). The y-axis plots the percentage of /iŋ/ word responses. Error bars present the mean and 95% HDI of posterior predictions of velar responses.

5. PERCEPTION: GOODNESS RATING

Only the four variants of target /iN/ minimal pairs were included in this task to abbreviate the experiment; no filler trials were included. In each trial, participants heard the stimulus while seeing a Chinese character and a continuum ranging from 1 to 7 in whole numbers. Listeners were asked to click on the rating that indicates how well the auditory stimulus matched the character, with 1 being the least and 7 being the most accurate. Participants started with four practice trials. The screen automatically advanced to the next trial in 5s. A 1-sec fixation reset pause was added between trials to avoid habitual clicking.

5.1. Materials

Same critical items were used as in Section 4.1. Each variant had two production tokens repeated twice with each appropriate Chinese character in each block for three randomized blocks (totaling 144 trials, 6 trials for each lexical target).

5.2. Goodness rating results

Similar to the categorization task, within-group comparisons were made. Four Bayesian mixed effect linear regression models were fitted separately for each dialect group and for when the displayed word ends with /n/ and /ŋ/ using the {brms} package [20] in R [21]. The outcome variable was the Rating and the predictor variable was the dummy-coded Auditory Stimuli Type ([in], [iŋ], [iɲ], [iəŋ]; reference level: [iŋ]). A by-subject random intercept with Stimuli Type as by-subject slope and a by-word random intercept were included. Samples were drawn from the posterior distribution using a four-chain Hamiltonian Monte-Carlo sampling (2000 iterations, 500 warm-ups). Weakly informative priors were used for all parameters. Figure 3 shows the group ratings for the four variant types with error bars indicating the 95% HDI range of posterior predictions.

The models provide evidence that [iəŋ] is rated as the best /iŋ/ exemplar by both BjM (β = 6.76, CrI = [0.45, 1.60], PD = 99.95%) and ShM group (β = 5.66, CrI = [0.02, 3.04], PD = 97.58%). BjM listeners rated [in] as the best /in/ exemplar (β = 6.61, CrI = [2.34, 3.82], PD = 100%), [iɲ] as a better match for /in/ than /iŋ/ (β = 5.22, CrI = [1.12, 2.22], PD = 100%) and [iŋ] as a better match for /iŋ/ than /in/ (β = 5.74, CrI = [4.60, 6.85], PD = 100%). The models show little evidence for difference across ShM listeners' rating for [in], [iɲ], and [iŋ] (PD < 75%) and the posterior predictions in Figure 3 show that ShM listeners consistently give higher /in/-word ratings to all these three variants.

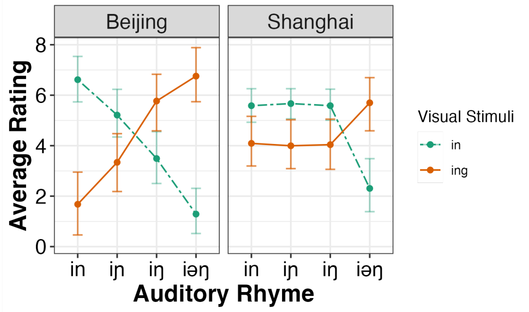

Figure 3: Goodness ratings for ShM and BjM with the y-axis showing the rating from 1 (worst) to 7 (best) for the stimuli types. The error bars indicate the 95% HDI of posterior predictions.

6. DISCUSSION AND CONCLUSION

In production, it is clear that BjM speakers contrast /iN/ and produce variants with no overlap. ShM speakers, at a group level, show a low yet existing degree of distinction of [iŋ]-[iəŋ] for /iN/ rhymes. This pattern could not have been observed when the investigation of ShM production was limited to a single nasal segment level, since the nasals in both variants are /ŋ/. The role of vowel for ShM suggests that some contrast at the rhyme level remains. This could indicate a shift in contrastive acoustic cues from nasal coda to vowel, but this requires further investigation.

Perceptually, both BjM and ShM listeners consider [iəŋ] as the best exemplar for /iŋ/. ShM listeners exhibit lower-to-no degree of distinction in lexical mappings for variants that only differ on the nasal segments – that is, at the group level there is not strong evidence that a particular nasal coda maps to a particular lexical item. However, BjM listeners still exhibit contrastive lexical mappings of variants that have no vowel quality difference ([in] vs. [iŋ]), which contradicts previous results [3, 16]. This suggests that the type of perceptual task (AXB task vs. lexical categorization task) and response (Pinyin label vs. Chinese character) elicit different patterns and conclusions. Lexical information can facilitate categorization [23], and the use of a single voice could have improved listener performance [24].

When comparing each group's perception to their own production, it is interesting to see that ShM users' exemplar for /iŋ/ does not match across perception and production. One possible explanation is ShM users' frequent exposure to BjM and Standard Mandarin via media and education. Due to the promotion of Standard Mandarin, speakers of all Mandarin varieties were taught and tested on the contrast in coda nasal through the Pinyin representation even when they do not adopt such distinction in production. Therefore, when presented with the auditory stimuli [iəŋ], which is "foreign" to ShM listeners, they could still map it to the canonical BjM productions of /iŋ/ words. This suggests a possibility of social factors and language exposure influencing listeners' perceptual exemplars even when such exemplars are not present in their production space.

To sum up, the current project revisits the distribution of Mandarin coda nasals in ShM and BjM users' perception and production. It provides insights into the necessity of considering the degree of distinction above sound segment level in both perception and production. This project focusing on group-level does not account for the large degree of individual variation. Follow-up studies investigating the individual perception-production (mis)alignment and potential generational differences are in progress.

7. ACKNOWLEDGMENTS

Thank you to Roger Lo for statistical guidance. Kathleen Currie Hall provided comments that improved the clarity of this work.

8. REFERENCES

[1] J. Weng, "What is mandarin? the social project of language standardization in early republican china," *The Journal of Asian Studies*, vol. 77, no. 3, pp. 611–633, 2018.

[2] L. Wang, J. Cui, and Y. Chen, "Wuxi speakers' production and perception of coda nasals in mandarin." in *INTERSPEECH*, 2018, pp. 2559–2562.

[3] M. Faytak, S. Liu, and M. Sundara, "Nasal coda neutralization in shanghai mandarin: Articulatory and perceptual evidence," *Laboratory Phonology*, vol. 11, no. 1, 2020.

[4] C. C. Lin, "Nasal endings of taiwan mandarin: Production, perception, and linguistic change," in *35th International Conference on Sino-Tibetan Languages and Linguistics, Arizona State University*, 2002.

[5] M. Luo, M. Sloos, and J. van de Weijer, "Perception and production of mandarin nasal codas by shanghainese speakers," *Chinese Accents and Accented Chinese CAAC2014*, p. 16, 2015.

[6] C. Chiu, Y. Lu, Y. Weng, S. Jin, W. Weng, and T. Yang, "Uncovering syllable-final nasal merging in taiwan mandarin: Ultrasonographic investigation of tongue postures and degrees of nasalization," in *Proceedings of the 19th International Congress of Phonetic Sciences*, 2019, pp. 398–402.

[7] J. Fon, J.-m. Hung, Y.-H. Huang, and H.-j. Hsu, "Dialectal variations on syllable-final nasal mergers in taiwan mandarin," *Language and Linguistics*, vol. 12, no. 2, pp. 273–311, 2011.

[8] H.-j. Hsu and J. K.-P. Tse, "Syllable-final nasal mergers in taiwan mandarin-leveled but puzzling," *Concentric: Studies in Linguistics*, vol. 33, no. 1, pp. 1–18, 2007.

[9] Y.-H. Lai, "Production of mandarin chinese nasal coda by l1 and l2 speakers of mandarin chinese," 华语文教学研究*[Huayu Wenjiaoxue Yanjiu]*, vol. 5, no. 1, pp. 155–180, 2008.

[10] Y. Guan, "Nasal coda realisation in speech production of shanghai mandarin," 2019.

[11] Y.-h. Lai, "Perceptual processing of mandarin nasals by l1 and l2 mandarin speakers," *Journal of psycholinguistic research*, vol. 41, no. 4, pp. 237–252, 2012.

[12] X. Mou, "Nasal codas in standard chinese: a study in the framework of the distinctive feature theory," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.

[13] Q. Ran, "汉语鼻音韵尾的实验研究[An experimental study of Mandarin nasal codas]," 南开语言学刊*[Nankai Yuyanxue Kan]*, no. 2, pp. 37–44, 2005.

[14] F. Shi, "北京话的元音格局[The Vowel Distribution of Beijing Hua]," 南开语言学刊*[Nankai Yuyanxue Kan]*, no. 1, pp. 30–36, 2002.

[15] Y. Chen and S. Guion-Anderson, "Perceptual confusabiltiy of word-final nasals in southern min and mandarin: Implications for coda nasal mergers in chinese." in *ICPhS*, 2011, pp. 464–467.

[16] W. Wu, "普通话鼻音韵尾的听辨实验研究[a perception discrimination study of mandarin nasal codas]," Master's thesis, 天津师范大学[Tianjin Normal University], 2019.

[17] K. J. Woods, M. H. Siegel, J. Traer, and J. H. McDermott, "Headphone screening to facilitate web-based auditory experiments," *Attention, Perception, & Psychophysics*, vol. 79, no. 7, pp. 2064–2072, 2017.

[18] X. Xie, R. M. Theodore, and E. B. Myers, "More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 43, no. 1, p. 206, 2017.

[19] A. L. Anwyl-Irvine, J. Massonnié, A. Flitton, N. Kirkham, and J. K. Evershed, "Gorilla in our midst: An online behavioral experiment builder," *Behavior research methods*, vol. 52, no. 1, pp. 388–407, 2020.

[20] P.-C. Bürkner, "Bayesian item response modeling in R with brms and Stan," *Journal of Statistical Software*, vol. 100, no. 5, pp. 1–54, 2021.

[21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[22] B. Nicenboim and S. Vasishth, "Statistical methods for linguistic research: Foundational ideas—part ii," *Language and Linguistics Compass*, vol. 10, no. 11, pp. 591–613, 2016.

[23] M. H. Davis and I. S. Johnsrude, "Hearing speech sounds: top-down influences on the interface between audition and speech perception," *Hearing research*, vol. 229, no. 1-2, pp. 132–147, 2007.

[24] C. E. Stilp and R. M. Theodore, "Talker normalization is mediated by structured indexical information," *Attention, Perception, & Psychophysics*, vol. 82, pp. 2237–2243, 2020.