Interspeech 2018
2-6 September 2018, Hyderabad

Interactions Between Vowels and Nasal Codas in Mandarin Speakers' Perception of Nasal Finals

Chong Cao, Wei Wei, Wei Wang, Yanlu Xie, Jinsong Zhang

Beijing Advanced Innovation Center for Language Resources Beijing Language and Culture University, Beijing, China

caochong2013@126.com, weiweihyy0901@163.com, 15300461@qq.com, xieyanlu@blcu.edu.cn, jinsong.zhang@blcu.edu.cn

Abstract

Previous studies have shown vowels' effect on the perception of nasal codas from the perspectives of acoustics and phonetics with evidence from Mandarin. While few studies investigated the processing interactions between vowels and nasal codas during the perceptual processing of nasal Finals. Using a speeded classification paradigm which requires participants to attend to one dimension selectively and to ignore other dimensions, we aimed to explore the processing interactions between vowels and nasal codas in the process of nasal Finals' perception. Results of the speeded classification experiment showed that there was a mutual and symmetrical pattern of interference effect between vowel and nasal coda dimensions. More specifically, participants were affected by irrelevant variation of the nasal coda dimension when they were attending to the vowel dimension and affected by irrelevant variation of the vowel dimension when they were attending to the nasal coda dimension. The extent of such mutual interference effect was equal. Results were discussed in terms of phonemic differences in acoustic properties and relative discriminability between vowels and nasal codas. The present study may be helpful to second language learners of Mandarin during their course of acquiring nasal Finals.

Index Terms: vowels, nasal codas, processing interactions

1. Introduction

The Final is an indispensable part in a syllable which can be composed of a vowel, a diphthong, or a triphthong and with an optional nasal coda in standard Mandarin. Finals ending with nasal codas are defined as nasal Finals [1]. There are two contrastive nasal codas in standard Mandarin, i.e., alveolar nasal /n/ and velar nasal /ŋ/. These two nasal codas combine with vowels to make different nasal Finals [2]. In the modern Chinese phonetics, there is a total of 39 Finals, with the proportion of nasal Finals standing at 41% (i.e., 16 nasal Finals). There are 405 basic monosyllables without considering the neutral tone and four lexical tones, among which 177 are composed of nasal Finals accounting for 44% [3]. With respect to Chinese characters, there are 1387 characters with nasal Finals which are responsible for 46% of the 3000 frequently-used characters. Thus, it is very important for listeners to correctly perceive nasal Finals during the process of speech communication.

A number of studies have looked into the perceptual cues of /n/ and /ŋ/ when combined with different vowels in the context of VC syllables. For instance, Zee showed the effect of the vowel quality on the identification of the post-vocalic nasal consonants (i.e., /n/, /ŋ/), /n/ and /ŋ/ were best identified after the vowel /a/ among the five vowels [4]. Malécot reported that the formant transitions between vowels and nasal consonants were important cues for place identification between /n/ and /ŋ/ [5]. Larkey et al. indicated that the second and the third formant (i.e., F2, F3) were the primary acoustic cues for the realization of categorical identification of synthetic nasal consonants [6].

The bulk of above researches concerning the perceptual cues of /n/ and /ŋ/ in VC syllables was mainly concentrated on English. It was assumed that nasal codas (i.e., /n/, /ŋ/) in standard Mandarin were different from nasal consonants in English [7]. With respect to Mandarin, a series of studies have investigated the perceptual cues of /n/ and /ŋ/ in VC syllables. For instance, some perceptual studies showed clear evidence for effects of vowel qualities on identification of /n/ and /ŋ/, more specifically, the more similar the phonetic values (i.e., formants) of the vowels were, the less distinctive that the two nasal codas were [8, 9], which was comparable to the result of [4]. Nasal Finals in Mandarin were divided into three parts: vowel nucleus, nasalized vowel and nasal coda [10]. These three parts were proved to have different contributions to identification of /n/ and /ŋ/. [11] found that nasal codas could only provide cues of the manner of articulation instead of the place of articulation by truncating nasal codas in nasal Finals' classification tests. [12] found that the nasalized vowel was the most significant cue for Chinese subjects to judge whether a syllable is alveolar nasal-ended (i.e., /n/) or velar nasal-ended (i.e., /ŋ/). Based on the results of [11, 12], [13] further revealed that F2 and F3 of the nasalized vowel were the primary acoustic cues for the realization of categorical perception for Chinese subjects through identification tasks with continua of nasal Finals.

It is noteworthy that aforementioned studies concerning the perceptual cues of /n/ and /ŋ/ in Mandarin have mainly demonstrated vowels' effect on identification of nasal codas. The interaction between vowels and nasal codas, to our knowledge, was less studied. It is thus unknown how vowels and nasal codas interact during the perceptual processing of nasal Finals. In order to fill this knowledge gap, the current study aimed to explore the processing interactions between vowel and nasal coda dimensions in the process of nasal Finals' perception. We used the Garner speeded classification paradigm which was proposed by [14-16]. This procedure was considered to be well suited to the purpose of the present study, in that it effectively reveals the type of dimensional interactions between multidimensional stimuli within a single signal. Garner speeded classification studies of Mandarin have estimated the interactions between segmental (i.e., consonant, vowel) and suprasegmental dimensions (i.e., tone) [17-24]. In

10.21437/Interspeech.2018-2025

addition, a variety of studies demonstrated the processing interactions between two kinds of segmental dimensions, i.e., consonants and vowels [24, 25].

The principal purpose of the present study was to tap further into the issue of interactions between vowels and nasal codas during the perceptual processing of Chinese nasal Finals. We aim to address two questions:

- What is the pattern of dimensional interference between vowels and nasal codas? Separable or integral?
- Whether the extent of the interference effect is equal or not for both dimensions (i.e., vowels and nasal codas)?

2. Method

2.1. Garner speeded classification paradigm

Participants in the Garner task identified the attribute of the target dimension which exists at two or more levels in following three conditions:

1. Single dimension (i.e. control group). Only the target dimension varies and the nontarget dimension is held constant.

2. Correlated dimension. Both the target dimension and nontarget dimension vary, and each attribute of one dimension occurs with only one of the attributes of the other dimension.

3. Orthogonal dimensions. Both the target dimension and nontarget dimension vary orthogonally, and each attribute on one dimension occurs with each attribute of the other dimension.

The Garner's task requires participants to pay attention to one dimension selectively, while trying to ignore other dimensions. It was assumed that if the two dimensions are perceptually separable, the classification speed reflected by reaction times (i.e., RTs) will be the same in all three conditions. If, on the other hand, the two dimensions are perceptually integral, the classification speed in the correlated condition will be faster to facilitate the discrimination (i.e., redundancy effect) and slower to interfere with the discrimination in the orthogonal condition (i.e., interference effect) [16].

2.2. Stimuli

Two pairs of syllables with Tone 1 in the context of a VC syllable, /ən/-/əŋ/ and /in/-/iŋ/ were selected as the target stimuli. They were chosen with two primary goals: first, the vowels of these two pairs do not undergo assimilation when combined with nasal codas, which is different from the vowel /a/, /a/ was assimilated as /ɑ/ when combined with the velar nasal /ŋ/ [1, 2]; second, the vowel in each pair of /Vn/-/Vŋ/ are phonetically the same according to their respective International Phonetic Alphabet (i.e., IPA), thus, each pair of syllables constitutes a minimal pair which differs only in nasal codas. Each syllable was approximately 550 ms in duration.

2.3. Procedure

We conducted a preliminary classification experiment to obtain some estimate of the relative discriminability of vowel and nasal coda dimensions. Subjects were asked to distinguish vowels and nasal codas as quickly as possible without sacrificing the response accuracy. Results showed that the selected stimuli were approximately equal in discriminability for the both dimensions.

The formal speeded classification experiment consisted of two tasks, with the target and nontarget dimensions: vowel target/nasal coda nontarget (i.e., V/N) and nasal coda target/vowel nontarget (i.e., N/V). In each task, subjects were asked to classify the stimuli according to the target dimension (i.e., vowel or nasal coda) and to press the corresponding response keys as rapidly and as accurately as possible. The keys were labeled /ə/ and /i/ for the vowel judgement task (i.e., V/N), and /n/ and /ŋ/ for the nasal coda judgement task (i.e., N/V). There were five conditions per task: two control conditions, two correlated conditions and one orthogonal condition. The order of the conditions in each task was fixed as in the experiment of [17]: control, correlated, orthogonal, correlated, control. Table 1 lists the specific stimuli which were displayed in the form of official Pinyin transcription in five conditions. In each control condition, subjects heard two stimuli in which only the values of the target dimension varied. In each correlated condition, subjects heard two stimuli in which the value of the target dimension occurred with only one of the values of the nontarget dimension. In each orthogonal condition, subjects heard four stimuli in which the values on the target dimension as well as the nontarget dimension were varied orthogonally. Each condition included 25 repetitions of the stimuli in random order. Thus, there were 50 stimuli in each control and correlated condition, and 100 stimuli in each orthogonal condition. There was a total of 300 stimuli in each task. Between conditions there was a 10 s pause. Total time of two tasks per subject ranged from 40-60 mins.

Table 1: *Stimuli used in five conditions.*

Condition	Task	
	V/N	N/V
Control	(en, in)	(in, ing)
Correlated	(en, ing)	(ing, en)
Orthogonal	(eng, in, en, ing)	(en, in, eng, ing)
Correlated	(eng, in)	(in, eng)
Control	(eng, ing)	(en, eng)

2.4. Subjects

Twenty subjects of native Mandarin speakers from Beijing Language and Culture University participated the experiment. The age of subjects ranged from 23 to 28 years old. None of them have language, hearing or speaking impairments.

3. Results

Response accuracy and mean RTs in correct response trials were computed for each subject, task and condition. The dimensional interference effect was evaluated by comparing the average RTs of the orthogonal condition with its corresponding control condition. Any significant differences would indicate that the identification of the target dimension was affected by the variation in the non-target dimension. The extent of the interference effect (i.e., dimensional integrality) was measured by RT_{diff} which was calculated by subtracting the RTs of the control condition from the RTs of the corresponding orthogonal condition [14-16]. In the following, we will discuss the response accuracy, dimensional interference and dimensional integrality in turn.

3.1. Response accuracy

The average response accuracy for each task and condition was computed. In the V/N task, participants averaged 99.67% correct classification across 5 conditions, ranging from the highest response accuracy 100% to the lowest response accuracy 98.33%. In the N/V task, participants averaged 99% correct classification across five conditions, ranging from the highest 100% to the lowest 92%. Generally, subjects performed the tasks quite accurately. Table 2 presents the detailed error percentages of each condition in each task. It can be seen from table 2 that the average overall error rate was less than 2% in each condition of each task.

Table 2: *Error percentages of each condition.*

Task	Condition		
	Control	Correlated	Orthogonal
V/N	0.15%	0.40%	0.45%
N/V	0.35%	1.20%	1.45%

3.2. Dimensional interference

The mean RTs and standard deviation were calculated, if any individual RT fell more than two standard deviations from the mean, it was considered an outlier and was excluded from analysis, the mean was recomputed until no outliers remained. In this way, variance across subjects, tasks and conditions was stabilized to some extent. Figure 1 shows the principal results for the RTs. Each column represents an average RTs of three experimental conditions: control, correlated and orthogonal. These means are based on correct responses only.

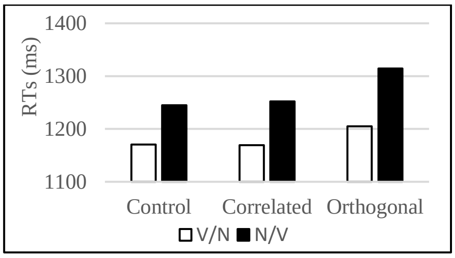

Figure 1: *Average RTs in each condition.*

It can be seen from figure 1 that RTs are slower in the nasal coda classification task than in the vowel classification task for each condition. A closer look into the vowel classification task suggests that RTs of the orthogonal condition are obviously slower than the corresponding control condition, with the correlated and control means not reliably different. A closer look into the nasal coda classification task suggests that both the orthogonal and correlated condition means are longer than the corresponding control means. The average values of RTs (in milliseconds) in each condition for each judgment task (i.e., V/N, N/V) are shown in Table 3.

Table 3: *The reaction times of each condition.*

Task	Condition		
	Control	Correlated	Orthogonal
V/N	1170.64	1169.41	1204.78
N/V	1244.62	1252.23	1314.32

To confirm the existence of the interference effect and redundancy effect, RTs in each orthogonal condition and correlated condition were compared with its corresponding control condition by using a paired *t* test, respectively. In the vowel judgement task (i.e., V/N), the RTs of the orthogonal condition were significantly slower than its corresponding control group at a $p<0.01$ level ($t=3.826$, $p=0.001$), which suggested that participants were affected by irrelevant variation in the nasal coda dimension when they were attending to the vowel dimension. Whereas no significant differences in RTs were obtained between the correlated condition and the control condition ($t=-0.209$, $p=0.837$), which suggested that redundant variation in the nasal coda dimension did not aid vowels classification. In the nasal coda judgement task (i.e., N/V), the RTs of the orthogonal condition were significantly slower than its corresponding control condition at a $p<0.01$ level ($t=3.979$, $p=0.001$), which suggested that participants were affected by irrelevant variation in the vowel dimension when they were attending to the nasal coda dimension. Whereas no significant differences were found between the correlated condition and the control condition ($t=0.462$, $p=0.649$), which suggests that redundant variation in the vowel dimension had no reliable effect on classification times of nasal codas.

In a word, both judgment tasks (i.e., V/N and N/V) suggested an identical pattern of results. Specifically, results demonstrated the presence of an interference effect, with an absence of a redundancy effect. Therefore, we could estimate that there was a mutual interference effect between vowel and nasal coda during the perceptual processing of nasal Finals.

3.3. Dimensional integrality

The above analysis of the significant interaction effect revealed a mutual interference effects between vowels and nasal codas. It was important to determine whether the extent of integrality was equal. That is, whether the degree of vowel interfering with nasal coda was the same as that of nasal coda interfering with vowel. If the interference effect is the same for one dimension as the other, the interference is mutual and symmetrical; if the interference effect is greater for one dimension than the other, the interference is mutual, but asymmetrical [16]. As we mentioned before, both symmetrical and asymmetrical effects were measured by RT_{diff} which was calculated by subtracting the RTs in the control condition from the RTs in the orthogonal condition. The mean RT_{diff} of two judgement tasks are shown in figure 2. Initial analysis revealed a longer RT_{diff} in N/V task compared with that in V/N task. Paired sample *t* tests taking the RT_{diff} as the dependent variable showed no significant differences ($t=1.791$, $p=0.089$), which indicated that the extent of integrality between vowel and nasal coda dimensions was equal.

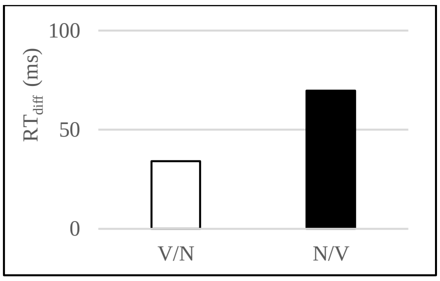

Figure 2: *RT_{diff} in each task.*

4. Discussion

The experiment was designed to investigate two research questions concerning the pattern of dimensional interference between vowels and nasal codas, as well as the extent of interference effect.

The result of the Garner speeded classification task showed that each orthogonal condition of each judgement task had a slower RTs than its corresponding control condition, with correlated condition and control condition means not significantly different for both judgment tasks. Such data demonstrated the presence of an interference effect, with an absence of a redundancy effect during the perceptual processing of nasal Finals in Mandarin. More specifically, when subjects were asked to make a judgment regarding the vowel, orthogonal variation in the nasal coda dimension significantly slowed RTs, whereas redundant variation in the nasal coda dimension had no reliable effect on classification times. When subjects were asked to make a judgment regarding the nasal coda, orthogonal variation in the vowel dimension significantly slowed RTs, whereas redundant variation in the vowel dimension had no reliable effect on classification times. Moreover, results of RT_{diff} for both two judgement tasks suggested that the extent of the interference effect between vowel and nasal coda dimensions was equal. In conclusion, a mutual and symmetrical interference effect between vowel and nasal coda was obtained during the perceptual processing of nasal Finals.

Prior studies concerning the processing interactions between vowel and consonant indicated that vowel dimensions interfered more with consonant's classification than the reverse [17]. They argued that the asymmetric interaction can be explained with the key acoustic properties of vowels and consonants. Vowels which are characterized by the first three formants have static acoustic cues, while consonants characterized by the rapidly changing bursts and formant transitions have dynamic acoustic cues [1]. Static and dynamic cues have different decay rates in auditory working memory [26]. Compared with static cues, dynamic cues decayed faster [27], therefore, the working memory representation of a vowel persisted more strongly than that of consonants. In this study, the symmetrical interaction between vowel and nasal coda can also be accounted for in terms of the relative degree of change over time (dynamic vs. static) in key acoustic properties of vowels and nasal codas. It has been well established that the first three steady-state formants are the primary acoustic cues for vowels [1], and the second and third formants of the vowel part are the key acoustic cues for nasal codas [12, 13]. Both the vowel and nasal coda were characterized by the static acoustic cues, therefore, the working memory representation of vowels should be the same as that of nasal codas.

A critical issue to be addressed in this study is the discriminability along two dimensions (i.e., vowel and nasal coda). The result of the preliminary experiment suggested that participants could distinguish vowels accurately and quickly as well as nasal codas in the recognition processing of nasal Finals. But, the formal experiment demonstrated that RTs of the control condition in the nasal coda judgement task was significantly slower than that in the vowel judgement task, which indicated that there was a discrepancy between the discriminability of vowels and nasal codas. When the values on the two dimensions did not differ in discriminability, mutual and symmetrical interference was observed. When the values on one dimension were more discriminable than they were on the other, a pattern of asymmetrical interference was observed. However, the result of the present study showed that the discriminability did not affect the pattern of interference effects. We infer that such differences in interference effect may account in part for the relatively static nature of their core acoustic properties, since both vowels and nasal codas are primarily characterized by formant frequencies. It is also necessary to carry out more comprehensive studies regarding nasal Finals' perception in the future.

5. Conclusion

Focusing on the processing interactions between vowel and nasal coda dimensions and using the Garner speeded classification task, the major finding of this study is that, Mandarin listeners showed a mutual and symmetrical interference effect between vowel and nasal coda during the perceptual processing of nasal Finals.

6. Acknowledgements

This work is supported by Advanced Innovation Center for Language Resource and Intelligence (KYR17005), Research Funds of State Language Commission (ZDI135-51), Wutong Innovation Platform of Beijing Language and Culture University (16PT05), and the Fundamental Research Funds for the Central Universities(18YJ030006). The corresponding author is Jinsong Zhang.

7. References

[1] H.Q. Bao and M. C. Lin, *Essentials of experimental phonetics*. Beijing: Beijing University Press, 1994.

[2] B. R. Huang and X. D. Liao, *Modern Chinese*. Beijing: Higher Education Press, 2010.

[3] Y. G. Dong, "Study on teaching Japanese students to learn Chinese nasal vowels," *Chinese teaching in the word*, no. 4, pp. 66-70,1997.

[4] E. Zee, "Effect of vowel quality on perception of post-vocalic nasal consonants in noise," *Journal of Phonetics*, vol. 9, no. 1, pp. 35-48, 1981.

[5] A. Malécot, "Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique," *Language*, vol. 32, pp. 274-284, 1956.

[6] L. S. Larkey, J. Wald and W. Strange, "Perception of synthetic nasal consonants in initial and final syllable position," *Perception & Psychophysics*, vol. 23, no. 4, pp. 299-312, 1978.

[7] X. Mou, *Nasal codas in standard Chinese—a study in the framework of the distinctive feature theory*, Massachusetts: Massachusetts Institute of Technology, 2006.

[8] Y. J. Wang, "An experimental study on the perception and production of nasal codas by Japanese Learners of Chinese Putonghua," *Chinese Teaching in the World*, no. 2, pp. 47-60, 2002.

[9] F. Wu, "An Experimental Study on the Perceptual Process of Mandarin Nasals by Ll and L2 Learners," *Linguistic Science*, vol. 15, no. 3, pp. 268-279, 2016.

[10] M. C. Lin and J. Z. Yan, "Coarticulation in the zero-initial syllable with nasal ending in Standard Chinese," *Applied Linguistics*, vol. 13, no. 1, pp. 12-20, 1994.

[11] H. Wang, *The perception of Japanese learners on Standard Chinese nasal finals*. Beijing: Beijing Language and Culture University, 2012.

[12] J. S. Zhang and Z. Y. Wang, "The influence of vowel segments on Japanese learners' perception of Chinese nasal codas," *Journal of Tsinghua University (Science and Technology)*, vol. 57, no. 2, pp.164-169, 2017

[13] Y. Li, Y. Xie, L. Feng, et al., "The perceptual cues for nasal finals in standard Chinese," in *ISCSLP 2016-10th International*

Symposium on Chinese Spoken Language Processing，*October 17-20, Tianjin, China, Proceedings*, 2016, pp.1-5.
[14] W. R. Garner, "The stimulus in information processing," *American Psychologist*, no. 25, pp. 350-358, 1970.
[15] W. R. Garner, *The processing of information and structure*. Potomac, MD: Erlbaum, 1974.
[16] W. R. Garner, "Interaction of stimulus dimensions in concept and choice processes," *Cognitive psychology*, vol. 8, no.1, pp. 98-123, 1976.
[17] B. H. Repp and H. B. Lin, "Integration of segmental and tonal information in speech perception: A cross‐linguistic study," *The Journal of the Acoustical Society of America*, vol. 87, no. S1, pp. S46-S46, 1990.
[18] J. L. Miller, "Interactions in processing segmental and suprasegmental features of speech," *Perception & Psychophysics*, vol. 24, no. 2, pp.175-180, 1978.
[19] T. D. Carrell, L. B. Smith, D. B. Pisoni, "Some perceptual dependencies in speeded classification of vowel color and pitch," *Perception & psychophysics*, vol. 29, no. 1, pp. 1-10, 1981.
[20] L. Lee and H. C. Nusbaum, "Interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese," *Perception & Psychophysics*, vol. 53, no.2, pp. 157-165, 1993.
[21] M. Lin and A. L. Francis, "Effects of language experience and expectations on attention to consonants and tones in English and Mandarin Chinese," *The Journal of the Acoustical Society of America*, vol. 136, no. 5, pp. 2827-2838, 2014.
[22] C. C. Wood, "Parallel processing of auditory and phonetic information in speech discrimination," *Perception & Psychophysics*, vol. 15, no. 3, pp. 501-508, 1974.
[23] C. C. Wood, "Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 1, no. 1, pp. 3-3, 1975.
[24] Y. Tong, A. L. Francis and J. T. Gandour, "Processing dependencies between segmental and suprasegmental features in Mandarin Chinese," *Language and Cognitive Processes*, vol. 23, no. 5, pp. 689-708, 2008.
[25] C. C. Wood and R. S. Day, "Failure of selective attention to phonetic segments in consonant-vowel syllables," *Perception & Psychophysics*, vol. 17, no. 4, pp. 346-350, 1975.
[26] D. B. Pisoni, "Auditory and phonetic memory codes in the discrimination of consonants and vowels," *Perception & psychophysics*, vol. 13, no. 2, pp. 253-260, 1973.
[27] D. Mirman, L. L. Holt and J. L. McClelland, "Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 1198-1207, 2004.