PHONETIC REALIZATIONS OF NEUTRAL TONES IN WUHAN CHINESE DISYLLABIC WORDS

Zhou Duanchi

Hitotsubashi University
duanchizhou@gmail.com

ABSTRACT

The purpose of this study is to clarify the realisation of the neutral tone (NT) phenomenon in Wuhan Chinese, one of the Southwestern Mandarin varieties, based on quantitative analysis using acoustic-phonetic methods. Specifically, we examine 1) how many types of NT exist in Wuhan Chinese, and 2) what factors determine the pitch value of NT syllables. According to previous studies based on impressionistic observations of the Wuhan tone system, there are two types of NT, realized with a pitch value of 5 (highest pitch) or 3 (mid pitch level). Contrary to previous studies, however, our quantitative analysis revealed that there are four types of NT in Wuhan Chinese and that the pitch value of NTs is determined by the underlying tones of both the first and second syllables.

Keywords: neutral tone, Wuhan Chinese, cluster analysis.

1. INTRODUCTION

In Wuhan Chinese, one of the sub-varieties of Southwestern Mandarin Chinese, there are four basic lexical tones in the citation form. These are referred to as T1, T2, T3, and T4, respectively. Figure 1 schematically shows the pitch contours of the four tones. Pitch values are expressed using the 5-level tone letters of Chao [2], where 1 indicates the low range, 2 the mid-low range, 3 the middle range, 4 the mid-high range, and 5 the high range.

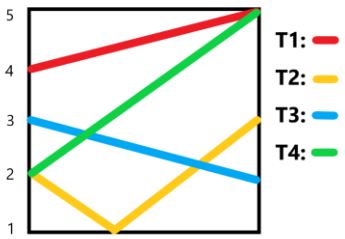

Figure 1: Schematic pitch contours for the four basic tones in Wuhan Chinese based on five-level tone letters.

In words with two or more syllables, the phenomenon of neutral tone (NT) may occur. This study defines NT as a phenomenon where the duration of a non-initial syllable is shortened and the number of tonal contrasts in that syllable is neutralised. In other Chinese dialects, there are rule-governed and non-rule-governed NT words [5,7]), and this holds true in Wuhan Chinese, too. In the former, NT is predictable grammatically, whereas it is lexically determined in the latter.

According to previous studies, there are two types of NT in Wuhan Chinese [3,6,9]: NT with pitch value 5 and with pitch value 3. Opinions differ regarding the factors that determine which of the two types of NT appears. Hou [3] and Zhu [10] argued that this factor is the tone of the first syllable. According to these studies, NT has a pitch value of 5 if the first syllable is {T1, T4}, whereas it has a pitch value of 3 if the tone of the first syllable is {T2, T3}. (Hereafter, "{A, B}" stands for "A or B"). In contrast, Li [6] states that the underlying tone of the second syllable itself determines the realisation of NT, and that if it is {T2, T3, T4}, then NT has a pitch value of 5. The abovementioned studies are based on impressionistic observations of Wuhan tones and provide no quantitative data, which may account for the absence of a consensus on the factors determining the pitch value of NTs. This study aims to clarify 1) how many types of NT exist in Wuhan Chinese, and 2) what factors determine the pitch value of NTs based on a quantitative analysis using acoustic-phonetic methods.

2. METHODS

2.1. Test words and speakers

A total of 300 disyllabic words, 150 that are expected to have an NT in the second syllable and 150 that do not, were used as test words. The criteria for the selection of each of the 150 words included the combinations of underlying tones of the first and second syllables. According to a widely accepted theory, Wuhan Chinese has four underlying tones; therefore, there are 4×4=16 possible combinations of underlying tones in disyllable words. However, Zhou [8] suggests that in Wuhan Chinese, the T2s derived from *Yangping* tones in Middle Chinese behave differently from those from *Rusheng* tones in the NT phenomenon. Therefore, in this study, T2 was divided into those derived from *Yangping* tone (henceforth,

T2) and those derived from *Rusheng* tone (henceforth, T2′). This gives 5×5=25 combinations of underlying tones. Providing six words for each combination yields 25×6=150 words of each category.

The speakers were two native speakers of Wuhan Chinese, F (born in 1961, female) and M (born in 1964, male), with no history of living abroad. Test words were presented to the speakers in a randomized order and the speakers were asked to read each word aloud in isolation. The produced speech data were measured using the speech analysis software Praat [1].

2.2. Measurements and statistical analysis

First, the test words were acoustically analysed. Based on a visual inspection of the waveforms and spectrograms, we identified the beginning and end of the rhyme of each syllable and measured its duration. After that, we divided the rhyme section of the second syllable into 20 sections and extracted the fundamental frequency (f_0) of that section.

Since there is a speaker-specific difference in the realisation of NT words (that is, some words are realised as NT words by one speaker but non-NT words by the other), the test words were classified as NT and non-NT words by means of a non-hierarchical cluster analysis that was performed for each speaker separately based on the rhyme duration of the first and second syllables of all test words. The cluster analysis classified the test words into two groups, and we considered the group with the shorter rhyme length to be the group with NT words.

Subsequently, we categorised the pitch values of NT words by performing a hierarchical clustering analysis for each speaker based on the f_0 of the rhyme section of the second syllable. As a method for measuring the distance between test words, we adopted the Dynamic Time Warping method (DTW), which is suitable for measuring time-series data, such as f_0 contours. Ward's method was used to measure the distance between the clusters. The mean value of the f_0 contour was then calculated for each obtained cluster.

Finally, we examined the factors that determine the pitch value of NT syllables by examining the correspondence between the underlying tone of the first and second syllables and the pitch value of NTs.

3. RESULTS

3.1. Classification of NT and non-NT words

Figure 2 shows the results of the non-hierarchical clustering analysis based on the duration of rhymes in the first and second syllables of all test words. The horizontal and vertical axes are the durations of the rhymes of the first and second syllables, respectively.

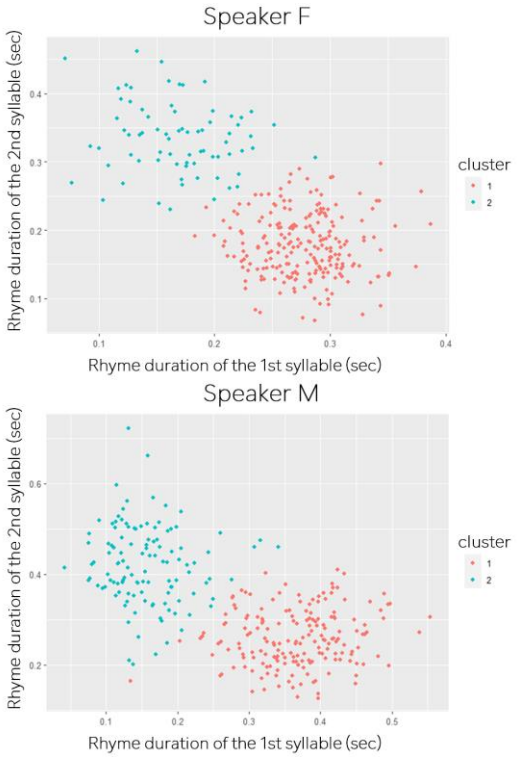

Figure 2: Results of non-hierarchical clustering analysis based on the duration of the rhymes of the 1st and 2nd syllables.

All words of both speakers were classified into Cluster 1 (red), with a longer first syllable and a shorter second syllable, and Cluster 2 (green), with a shorter first syllable and a longer second syllable. Words classified into Cluster 1 (red) were defined as NT words. The results of the non-hierarchical clustering analyses are almost consistent (with a match rate of more than 90%) with the classification based on the auditory impression of the author, who is a native speaker.

3.2. Classification of pitch values of NTs

Four interpretable clusters were extracted by the hierarchical clustering analysis based on the f_0 of the second-syllable rhyme of the NT words for each speaker. They are labelled A, B, C, and D. The max and min on the vertical axis are roughly the upper and lower limits of the pitch range of each speaker so that the gender differences are minimized.

To examine the pitch values of NTs, the mean f_0 of the rhymes in the second syllable was obtained for each cluster. Figure 3 illustrates the results for speaker F. The solid line indicates the mean f_0 for all words in each cluster, and the dotted lines indicate the mean f_0 obtained for each underlying tone of the second syllable in that cluster. In cluster A, the mean f_0 is located in the lowest region in the pitch range. In cluster B, the mean f_0 is located in the highest region.

In cluster C, the mean f_o is located in the middle region. In Cluster D, the mean f_o rises from the low to high regions. We refer to these pitch values of NTs respectively as L-NT, H-NT, M-NT, and R-NT.

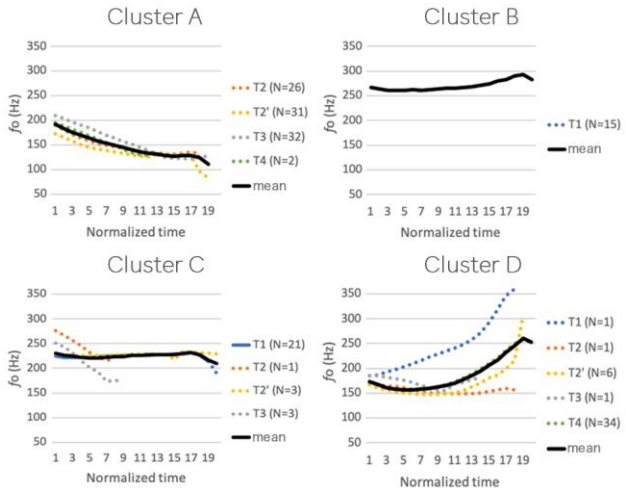

Figure 3: Mean f_o of the rhyme in the second syllable (Speaker F).

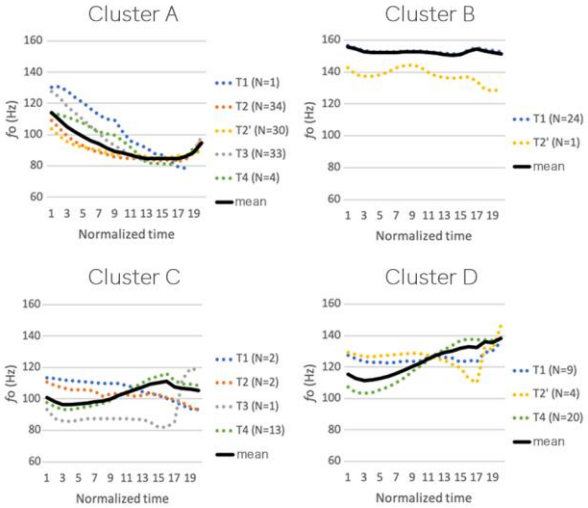

Figure 4: Mean f_o of the rhyme in the second syllable (Speaker M).

Figure 4 shows the results for speaker M. Similar to speaker F, L-NT appears in cluster A, and H-NT appears in cluster B. However, as far as the overall average value (solid line) is examined, both clusters C and D exhibit R-NT. The difference between the two clusters is in their pitch range; that is, the mean f_o is lower in cluster C than in cluster B. In contrast, examination of the mean f_o for each underlying tone of the second syllable (dotted line) reveals that M-NT and R-NT coexist in clusters C and D. Importantly, only words with an underlying T4 on the second syllable exhibit R-NT. Thus, as far as speaker M is concerned, f_o-based cluster analysis could not separate M-NT from R-NT. This was likely due to clustering based on differences in the overall pitch range rather than differences in f_o contour shape (rising or flat). The unstable phonation at the end of many of speaker M's words might have contributed to this failure.

3.3. Correspondence between underlying tones and NTs

To determine the factors conditioning the realisation of NTs, we examined the correspondence between each cluster and the underlying tone of its words.

Table 1 shows the distribution of the underlying tone of the second syllable (NT syllable) in each cluster for Speaker F. Almost all words with underlying {T2, T3} in the second syllable belong to cluster A (L-NT). Almost all words with an underlying T4 belonged to cluster D (R-NT). Words with an underlying T1 on the second syllable belong to clusters B (H-NT) and C (M-NT).

Since the realisation of NTs cannot be predicted from the underlying tone of the second syllable, we examined the underlying tone of the first syllable, finding first that all but 2 of the 15 words belonging to cluster B (H-NT) have a first syllable with an underlying T1 (5 words) or T4 (8 words), and second that all but 3 of the 21 words belonging to cluster C (M-NT) have a first syllable with an underlying T2 (7 words), T2′ (6 words), or T3 (5 words).

Since most of the words with underlying T2′ on the second syllable (31 words out of 40) belong to cluster A (L-NT), it seems reasonable to assume that they behave identically to words with T2, although the fact that a few T2′ words belong to clusters C (M-NT) and D (R-NT) deserves attention.

	T1	T2	T2'	T3	T4	Total
Cluster A （L-NT）	0	**26**	**31**	**32**	2	91
Cluster B （H-NT）	**15**	0	0	0	0	15
Cluster C （M-NT）	**21**	1	3	3	0	28
Cluster D （R-NT）	1	1	6	1	**34**	43
Total	37	28	40	36	36	177

Table 1: Breakdown of the underlying tone of the second syllable for each cluster (Speaker F).

Table 2 shows the results for Speaker M. Similar to Speaker F, almost all words with underlying {T2, T3} belong to Cluster A (L-NT). Most of the words with an underlying T1 (24 out of 36 words) belonged to cluster B (H-NT), whereas 11 words belonged to cluster C or D. As shown in Section 3.2, although M-NT and R-NT coexist in clusters C and D, words with underlying T1 on the second syllable are always realised as M-NT. Unlike speaker F, however, it is not possible to predict whether H-NT or M-NT will appear from the underlying tone of the first syllable.

As for those words with an underlying tone of T4, most belong to either clusters C or D, in which M-NT and R-NT coexist. However, words with an underlying T4 on the second syllable are only realised as R-NT (Figure 3). As in the case of T1, we cannot predict whether words with an underlying T4 belong to Cluster C (lower pitch range) or Cluster D (higher range) from the underlying tone of the first syllable. Thus, unlike the case for speaker F, we cannot confirm the involvement of the first syllable in the realisation of the NT.

Most of the words with underlying T2′ on the second syllable (30 out of 35 words) belonged to cluster A (L-NT). This leads to the assumption that they exhibit the same behaviour as words with T2, although the fact that some of them exhibit H-NT or M-NT may deserve attention.

	T1	T2	T2'	T3	T4	Total
Cluster A (L-NT)	1	**34**	**30**	**33**	4	102
Cluster B (H-NT)	**24**	0	1	0	0	25
Cluster C (M-NT, R-NT)	2	2	0	1	**13**	18
Cluster D (M-NT, R -NT)	9	0	4	0	**20**	33
Total	36	36	35	34	37	178

Table 2: Breakdown of the underlying tone of the second syllable for each cluster (Speaker M).

4. DISCUSSION

The results showed that there are four types of NTs in Wuhan Chinese, namely L-NT, H-NT, R-NT, and M-NT, and revealed that the underlying tones of both the first and second syllables determine which of the four types of NT is realized. The results can be summarised as (1).

(1) a. {T1, T2, T3, T4} + {T2, T3} → L-NT
b. {T1, T2, T3, T4} + T4 → R-NT
c. {T1, T4} + T1 → H-NT
d. {T2, T3} + T1 → M-NT

If the underlying tone of the second syllable is {T2, T3, T4} (1ab), then the underlying tone of the first syllable does not contribute to the realisation of NTs. On the other hand, if the underlying tone of the second syllable is T1 (1 cd), then the underlying tone of the first syllable contributes to NT realisation, although conditioning by the first syllable was not confirmed by the results of speaker M.

The NT phenomena make it possible to propose that in Wuhan Chinese, T2 and T3 constitute one natural class, whereas T1 and T4 form the other class. Each class can be considered to share its own distinctive tonal feature because the former has a low pitch that does not exceed the middle range in a citation form, whereas the latter has a high pitch above the middle range.

5. CONCLUSION

Wuhan Chinese has hitherto been described as having two types of NT, but the factors that determine the realisation of NTs have remained controversial [3,5,7]. The quantitative analyses of the present study based on acoustic-phonetic methods revealed that, in contrast to previous studies, there are four types of NT, and the underlying tones of both the first syllable and second syllables contribute to the realisation of NTs. Because the contribution of the first syllable's underlying tones was not detected for one of the two speakers, further studies with more participants should be conducted in the future. Our ongoing analysis of other speakers is yielding preferable results. At any rate, the previous description that there are only two types of NTs in Wuhan Chinese is not supported by the quantitative analysis of this study.

ACKNOWLEDGEMENTS

This study was supported by the NINJAL collaborative project "Empirical Study on the Intonational Diversity in Japanese and Ryukyuan Dialects" (project leader: Yosuke Igarashi).

6. REFERENCES

[1] Boersma, P., Weenink, D. 2011. Praat: Doing phonetics by computer [Computer program]. Version 5.2.16, from http://www.praat.org/.
[2] Chao, Y. R. 1930. A system of tone letters. *Le Maître Phonétique* 45, 24–27.
[3] Hou, J. Y. 1997. *Wuhanhua Yindang*. Shanghai: Shanghai Jiaoyu Chuban.
[4] Institute of Linguistics CASS. 2003. *Hanyu Fangyan Ciyu Diaocha Tiaomubiao. Fangyan*, 2003-01.
[5] Jing, S. 2022. *Xiandai Hanyu Qingsheng Dongtai Yanjiu*. Beijing: Minzu Chubanshe.
[6] Li, M. X. 2004. Neutral tones in disyllabic sequences across Chinese dialects: An OT account. Master's thesis, Tianjin Normal University.
[7] Lu, Y. Z. 2001. *Qingsheng he Erhua*, Beijing: Shangwu Yinshuguan.
[8] Zhou, D. C. 2021. Wuhan Fangyan Shuangyinjie Qingsheng De Shengdiao Tixi Yu Zhonggu Yinyun, *Proceedings of the 12th International Conference on Middle Chinese Studies*, 568–573.
[9] Zhu, J. S. 1992. *Wuhan Fangyan Yanjiu*. Wuhan: Wuhan Chubanshe.
[10] Zhu, J. S. 2017. *Wuhan Fangyan Cidian*. Wuhan: Chongwen Shuju.