TONAL TARGET AND PEAK DELAY IN MANDARIN NEUTRAL TONE

Aijun Li[1], *Zhiqiang, Li*[2], *Gan Huang*[3], *Liang Zhang*[1]
[1]Institute of Linguistics, Chinese Academy of Social Sciences, China
[2]Department of Modern and Classical Languages, University of San Francisco
[3]School of Chinese Language and Literature, Beijing International Studies University, China
liaj@cass.org.cn; zqli@usfca.edu; huanggan666@yeah.net; liangzhang1025@126.com

ABSTRACT

We examined the tonal target of the neutral tone syllable, F0 peak delay and the F0 preplanning process in production of Mandarin neutral tone by manipulating the number of neutral tone syllables and the preceding tonal contexts. The results showed that 1) the tonal target of neutral tone was L; 2) its realization was greatly influenced by the number of neutral tone syllables, as well as the prosodic structure; 3) the F0 pattern of the neutral tone depended on the tonal target of the proceeding non-neutral syllable. Specifically, the interpolation rule realized between the end position of F0 in T1 and T4, or the Peak Delay in T2 and T3, and the target position of neutral tone; and 4) with the increasing number of neutral syllables, the initial F0 in the prosodic unit also increased accordingly, indicating that our pitch preplanning ability was closely related to the prosodic structure.

***Index Terms* — neutral tone, F0 peak delay, tonal target, preplanning**

1. INTRODUCTION

Previous studies on Mandarin neutral tone have examined it from a wide swath of approaches. Some are lexicological [1], syntactic [2, 3, 4], and phonological [5, 6, 7, 8], while others have focused on its acoustic realization [9, 3, 10, 11, 12, 13] and perceptual correlates [4, 14, 15, 16, 17, 18].

Li [17] examines the realization of disyllabic neutral tone words as well as their non-neutral tone counterparts in five different information structures. Figure 1 illustrates the F0 patterns of non-neutral tone vs. neutral tone words in isolation, where the tonal contour of the neutral tone is dependent on the preceding non-neutral tone, showing a mid-falling tone after Tone 1(high, HH), a high falling tone after Tone 2 (rising, LH), a mid-tone after tone 3 (low dipping, LL(H), where H is a floating tone) and a low falling tone after Tone 4 (falling, HL), shown in Fig. 1.

Chen and Xu [13] indicates that neutral-tone syllables do have a target that is independent of the surrounding tones (as shown in Fig. 2), which is likely to be static and mid. Neutral tone differs from the other lexical tones in Mandarin

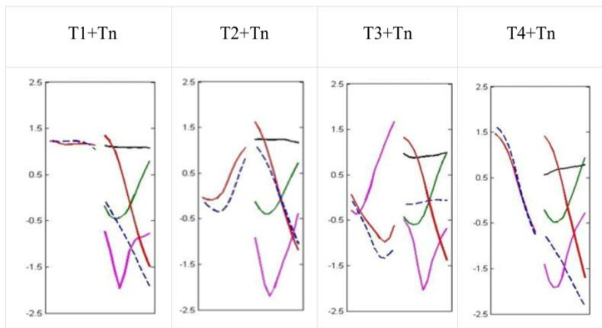


Fig 1. F0 contours of non-neutral tone (solid lines) vs. neutral tone words (dotted lines) grouped by the first syllable's tone

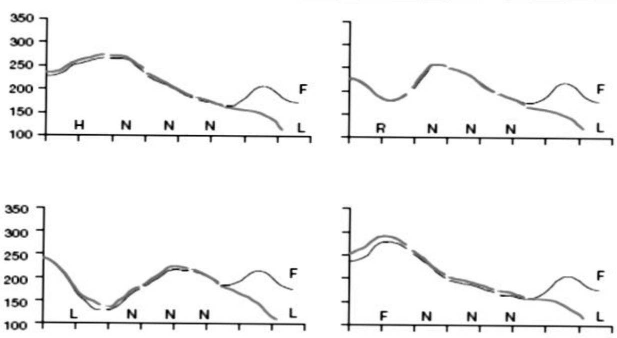


Fig 2. Neutral tone has a mid and static target tone in different contexts (after Chen and Xu, 2006)

Chinese not only in terms of its mid target, but also in terms of the weak articulatory strength with which this target is implemented.

Li [8] provides a preliminary study of F0 patterns in multiple neutral-tone contexts with a constraint-based analysis of tone target realization model.

Sun [20] shows that the neutral tone is characterized by [L] tonal feature. The neutral tone does not lose its low feature and realizes its [L] if there is not any impact from the adjacent tonal feature.

By manipulating the context, number of neutral tone syllables and tones of the proceeding non-neutral syllables, this study aims at examining how tonal target of neutral tone syllables, especially the one in the final position in a prosodic domain, and whether preplanning of F0 can be observed in the production of neutral tone.

2. ACOUSTIC EXPERIMENT

2.1. Participants

Thirty undergraduate and graduate students (aged between 18 and 25, with gender balanced) were recruited in the recording experiment for a small stipend. They were native speakers of Mandarin Chinese without a diagnosed reading or hearing disability.

2.2. Design and Material

The reading materials were classified into two groups, nouns or noun phrases in isolation and in carrier sentences. A 3*4 (number of neutral tone syllables * proceeding tones) within-participants design was employed. Table 1 shows the nouns or noun phrases with different number of neutral tone syllables in four tones, in which S stands for the four normal stressed tones in Mandarin, N for the neutral tone syllable, and the number indicates the position of each neutral tone syllable in the sequence. We used the same carrier sentences as Sun [20], which literally means "To send that bowl of flavor to SN1/SN1N2/SN1N2N3". Besides the same NP that used in the isolation condition, one more type of neutral tone syllable was added in the carrier sentence, ending with the aspectual marker "le" which is said in neutral tone.

Table 1. Target neutral tone syllables in isolation and in carrier sentences

Context	Tones	SN1	SN1N2	SN1N2N3
Isolation	T1	妈妈	妈妈们 妈妈的	妈妈们的
	T2	爷爷	爷爷们 爷爷的	爷爷们的
	T3	奶奶	奶奶们 奶奶的	奶奶们的
	T4	妹妹	妹妹们 妹妹的	妹妹们的
In carrier sentences	T1	妈妈 妈了	妈妈们 妈妈的 妈妈了	妈妈们的 妈妈们了
	T2	爷爷 爷了	爷爷们 爷爷的 爷爷了	爷爷们的 爷爷们了
	T3	奶奶 奶了	奶奶们 奶奶的 奶奶了	奶奶们的 奶奶们了
	T4	妹妹 妹了	妹妹们 妹妹的 妹妹了	妹妹们的 妹妹们了

2.3. Procedure

The nouns or noun phrases in isolation and in carrier sentences were divided into three groups randomly; each was read by ten speakers with gender balanced. The subjects read the materials in a sound-proof room. With a sampling rate at 16KHz, and 16bit quantitative resolution, the obtained recordings were controlled by a professional lab assistant to make sure they were in good quality. Finally, we recorded 160 files in isolation (40 in SN1, 80 in SN1N2, 40 in SN1N2N3), and 280 files in carrier sentences (80 in SN1, 120 in SN1N2, and 80 in SN1N2N3).

2.4 Data Analysis

All sound files were first automatically processed by a segmentation program to generate word-level, spelling-level and phone-level transcriptions for the recording, which can be read by Praat. Then careful manual correction and examination were taken to make sure the segmental boundaries were accurately marked. F0 contours of each recording were modified manually for erroneous pitch extraction. After annotation, F0 of each voiced phone was extracted in ten points at equal proportional intervals to normalize duration by a Praat script, with creaky voice excluded.

In order to eliminate speakers' individual differences, the F0 data of each speaker were transferred into Z-score in the formula:

$$F0z = (F0(target) - F0(mean))/ F0 (sd).$$

Where F0(mean) and F0 (sd) are the mean and the standard deviation of F0 data for each individual speaker.

3. RESULTS

In this part, we first present the results on the F0 contours of neutral tone syllables in different contexts, followed by the statistical analysis of the impact of the number of neutral tone syllables and tones of S, and the pitch resetting and preplanning at the beginning of the prosodic unit.

3.1. Tonal pattern and tonal target

3.1.1. SN1

Figure 3 (a) demonstrates that in disyllabic words with neutral tone, the F0 contour of the neutral tone syllable is dependent on the proceeding non-neutral tone, which is consistent with previous research [19]. Peak delay is observed when T2 (LH) and T3 (LLH) are involved, in which the H target is realized in the immediately following neutral tone syllable (N1). A consistent pattern of F0 contour of neutral tone syllables is also found in 3(b) in the carrier sentences. The aspectual marker shows a similar pattern as NP neutral-tone tones, expect for T2, in that NP (爷爷) shows a higher target point, which is more stressed.

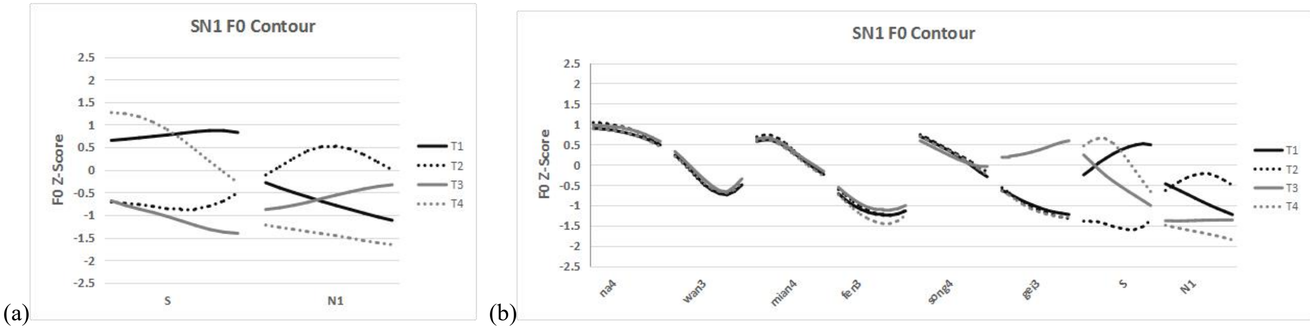


Fig 3. Mean and normalized F0 contours of SN1 in isolation (a) and in carrier sentences (b)

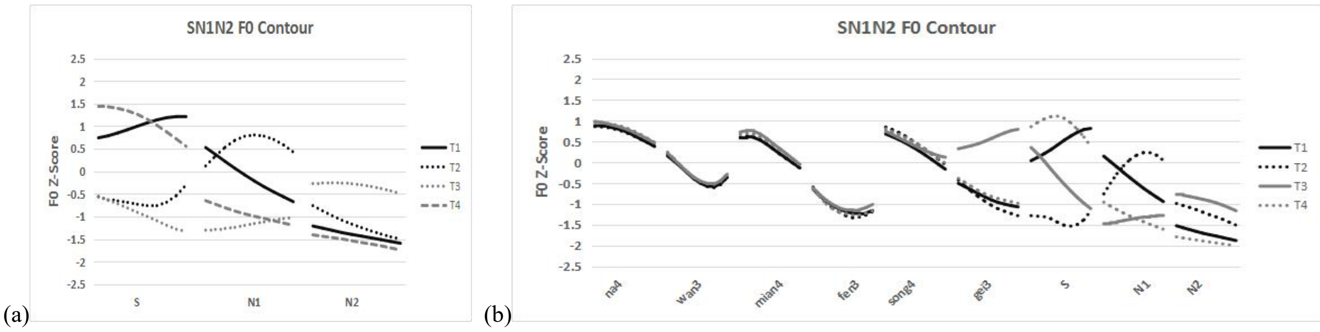


Fig 4. Mean and normalized F0 contours of SN1N2 in isolation (a) and in carrier sentences (b)

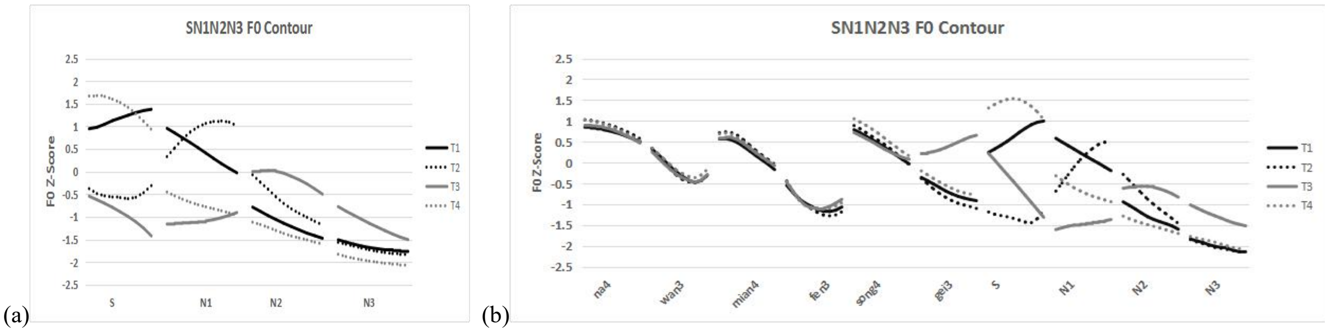


Fig 5. Mean and normalized F0 contours of SN1N2 in isolation (a) and in carrier sentences (b)

3.1.2. SN1N2

Figure 4 (a) shows that with two neutral tone syllables, their final target depends on the tone of the S syllable too, in that the final neutral tone syllable reaches a L target following T1, T2 and T4 and a M target following T3. However, different realization processes are observed. In the case of S being T1 and T4, the surface F0 on N1N2 seems to be interpolated from the ending high F0 to the final L target. Peak delay, though realized differently, is observed in the sequences headed by T2 and T3. When S is T2 (LH), its H target spreads to N1, and from there the F0 contour interpolates to the L target on N2. When S is T3 (LLH), its H target emerges on N2, phonetically realized as Mid. The same pattern is found in carrier sentences in Figure 4(b).

3.1.3. SN1N2N3

Figure 5 (a) illustrates that as the number of neutral tone syllables increases, the tonal target of the final neutral tone syllable tends to be appropriate a canonical L target aligned with the end of the syllable, which is also the end of the prosodic domain. As in SN1N2, the F0 contour on N1N2N3 after T1 and T4 is gradually declining to the final L target, while the peak delay happens in sequences headed by T2 and T3. Similar patterns are observed in carrier sentences with the end position of neutral tone syllable following T3 higher than other three tones (see Figure 5 (b)).

To conclude, in the context of multiple neutral tone syllables, the L target appeared only on the final syllable. The other non-final neutral tone syllables were not aligned with any clearly identifiable tonal targets. Rather, their surface F0 seemed to be interpolated from where the preceding non-neutral tone ends to the final L target. This is the case when the preceding tones were T1 (HH) and T4 (HL). When the proceeding syllables were in T2 (LH) and T3 (LL(H)), the H target was found to spread to the following neutral tone, leading to F0 peak delay. In the case of T2, the H target invariably surfaced on the immediately following neutral tone syllable (N1), regardless of the number of syllables in the sequence. In the case of T3, the H target was often realized on the next neutral tone syllable in SN1, but it could happen on the second or even on the third neutral tone syllable as in SN1N2 and SN1N2N3.

3.2. The statistical analysis on the neutral tone target

In order to examine the interactive impact of the designed contexts on the tonal target of the final neutral tone syllable (i.e. the final F0 value of the last neutral tone), General Linear Mixed Model (GLMM) was adopted, which suits best to the categorical data in the current study (fixed effects factors are number of neutral tones and the proceeding S tones; random effects factors are sentence ID and speaker ID). Further, One-Way ANOVA was adopted to examine the specific within-group differences in different numbers of neutral tone syllables. Repeated Measure was used in analyzing the difference of SN1N2N3 in isolation and in carrier sentences.

3.2.1. GLMM analysis

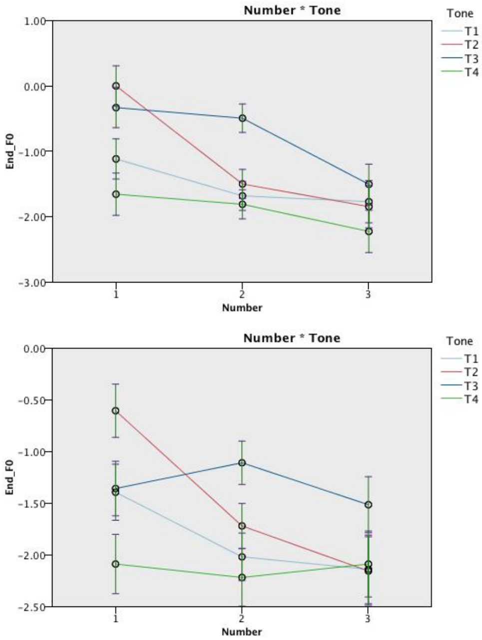
GLMM analysis results are showed in Figure 6 for isolation and carrier sentences respectively. Number, Tone and their Interaction are significant to the tonal target of neutral tone syllables both in isolation (Number, F (2, 141) =44.709, *p*<0.001; Tone, F (3, 141) =33.713, *p*<0.001; Interaction, F (6, 141) =7.287, *p*<0.001) and in carrier sentences (Number, F (2, 216) =18.699, *p*<0.001; Tone, F (3, 216) =20.051, *p*<0.001; Interaction, F (6, 216) = 0.194, *p*<0.001).

3.2.2. ANOVA analysis

In order to investigate the differences of neutral tone position following four tones in different number of neutral tone syllables, ANOVA analysis was adopted.

First, the results in isolation are reported. Please refer to Fig.6 (a). 1) In SN1, the result shows significant difference in proceeding tones (F (3, 35) =16.162, *p*<0.001), in which tone starting at L position (T2, T3) and starting at H position (T1, T4) significantly differ from the target position of the neutral tone syllable. 2) In SN1N2, the result shows a significant difference in tones (F (3,71) =43.983, *p*<0.001)), in which neutral tone following T3 is significantly higher than those following the other three tones. 3) In SN1N2N3, no significant difference is found in tones (F (2, 32) =2.515, *p*=0.076).

Accordingly, the ANOVA analysis in carrier sentences was conducted. Please refer to Fig.6 (b). 1) In SN1, the result shows significant difference in proceeding tones (F (3, 63) =15.871, *p*<0.001). T1 significantly differs from T2 (*p*=0.02) and T4 (*p*=0.028), and T2 significantly differs from T3 (*p*<0.001) and T4 (*p*<0.001). 2) In SN1N2, the result shows a significant difference in tones (F (3, 88) =22.230, *p*<0.001), in which the neutral tone following T3 is significantly higher than those following the other three tones. 3) In SN1N2N3, the result shows a significant difference in tones (F (3, 51) =8.552, *p*<0.001), in which the neutral tone following T3 is significantly higher than those following the other three tones.



(a)

(b)

Fig 6. The interaction between the number of neutral tone syllables and the proceeding tones for the target position F0 in isolation (a) and in carrier sentences (b)

It seems to be true from the ANOVA analysis that both in isolation and in carrier sentences, with an increase in the number of neutral tone syllable, the impact of the proceeding tones on the F0 scaling of the tonal target on the last neutral tone syllable is diminishing. In the condition of three neutral tone syllables, the end position is reliably aligned with a L target, despite T3 in carrier sentences shows a difference with *p*-value between 0.01 and 0.05.

3.2.3. Repeated measure

In order to examine the effect of the proceeding tone on the final targets in SN1N2N3 in isolation and in carrier sentences, we classified the four tones into two groups, T3 and the other three tones. A 2*2 (isolation/sentences * tone groups) condition was employed.

The repeated measurement shows isolation and carrier sentence conditions do not significantly differ at the end positions (F (1, 9) =0.037, *p*=0.851), while Tone groups do (F (1,9) =33.497, *p*<0.001), with T3 significantly higher than the other groups. Their interaction is also significant (F (1, 9) =15.055, *p*=0.004)).

T-test was employed to investigate their interplay. The results show that Tone groups significantly influence the target position both in isolation (t=-2.28, *p*=0.049) and in carrier sentences (t=3.937, *p*=0.001), while the target position is not significantly influenced by isolation or sentence conditions (*p*>0.05).

To summarize, the tonal target in the multiple neutral-tone context is realized consistently as the lowest position in the prosodic domain, thus carrying a Low tone feature.

3.3. F0 preplanning

We observe that the F0 of S is scaled higher as the number of neutral tone increase, as one can see in Fig. 3-5, a phenomenon called "speech preplanning" in Carlos' [21]. To examine the pitch preplanning effect, GLMM analysis was adopted in analyzing the impact of the number of neutral tone syllables and S tones on the initial F0 scaling of S, and the two more prosodic positions for carrier sentences.

3.3.1. Initial F0 in isolation

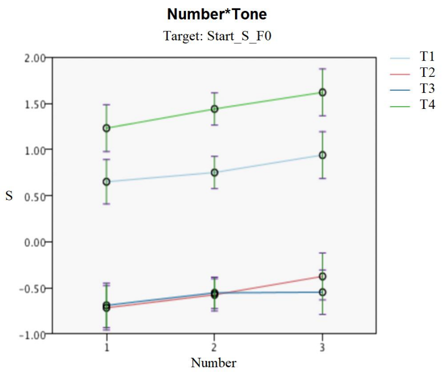


Fig 7. The interaction between number of neutral tone syllables and proceeding tones for the initial position of F0 in S syllable in isolation

GLMM shows that with the increased number of neutral tone syllables, the initial F0 climbs up in the phrase (F (2,141) =5.375, *p*=0.006)). It demonstrates a significant F0 preplanning in the initial F0 position in an intonation phrase.

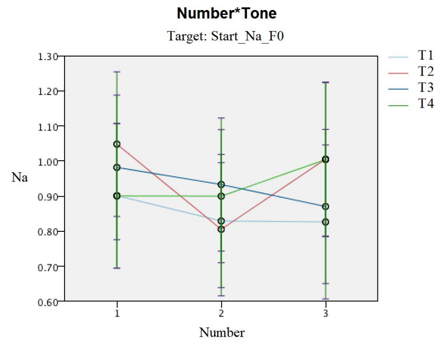
3.3.2. Initial F0 at different boundaries in sentences

We also checked initial F0 values in different prosody units: initial F0 of "na4" in an intonational phrase, initial F0 of "song4" in the second prosodic phrase and initial F0 of S in the prosody words containing N. Fig.8 presents the GLLM results.

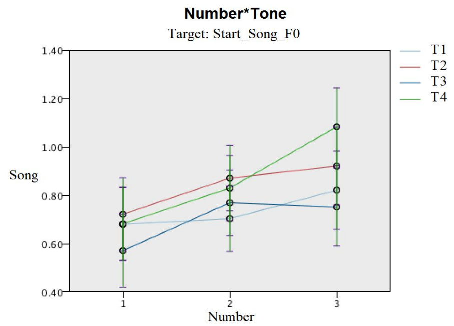
1) In the sentence-initial position in Fig. 8 (a), with the increase in number of neutral tone syllables or the variety of S' tones, there is no significant difference shown.

2) Unlike the sentence-initial F0, the initial F0 in the second prosodic phrase is more sensitive to the tones of the S syllable (F (3, 216) =4.219, *p*=0.006), especially in SN1N2N3, indicating that the initial tone of the prosodic word has anticipatory coarticulation effect to its previous prosodic words.

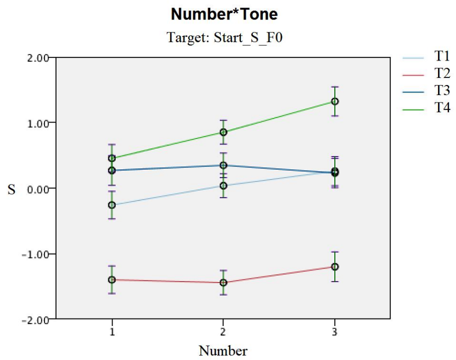
3) For the initial S syllable, the number (F (2, 210) =10.546, *p*<0.001)) and tone (F (3, 210) =180.815, *p*<0.001)) play significant roles separately and collectively (F (6, 210) =3.777, *p*=0.001). When T3 was excluded from the other three tones, we found that the longer the prosodic word, the higher the F0 reset in the initial position of the prosodic word, indicating that the pitch preplanning process is related to the prosodic structure.



(a) initial F0 of the sentence (IP)



(b) initial F0 of the second prosodic phrase



(c) initial F0 of syllable S

Fig 8. The interaction between number of neutral tone syllables and the proceeding tones for initial F0 values in sentence (a), second prosodic phrase (b), and S (c)

4. CONCLUSION

We examined the tonal target of the neutral tone syllable, F0 peak delay and the F0 preplanning process in production of Mandarin neutral tone by manipulating the number of neutral tone syllables and the preceding tonal contexts. Based on the results, we conclude that the tonal target of neutral tone is L, and its realization is greatly influenced by the number of neutral tone syllables, as well as the prosodic structure. The F0 pattern of the neutral tone depends on the tonal target of the proceeding non-neutral syllable. In the context of multiple neutral tone syllables, the L target appeared only on the final syllable. The other non-final neutral tone syllables were not aligned with any clearly identifiable tonal targets. Rather, their surface F0 seemed to be interpolated from where the preceding non-neutral tone ends to the final L target. This is the case when the preceding tones were T1 (HH) and T4 (HL). When the proceeding syllables are in T2 (LH) and T3 (LL(H)), the H target is found to spread to the following neutral tone, leading to F0 peak delay. In the case of T2, the H target invariably surfaces on the immediately following neutral tone syllable (N1), regardless of the number of syllables in the sequence. In the case of T3, the H target is often realized on the next neutral tone syllable in SN1, but it could happen on the second or even on the third neutral tone syllable as in SN1N2 and SN1N2N3. This conclusion supports the phonological analysis in Li Z.'s that the L target of the neutral tone syllable is aligned with the right edge of the prosodic domain consisting multiple neutral tone syllables [8]. The other part of the F0 contour is computed by way of F0 interpolation. Furthermore, with the increase in the number of neutral syllables, the initial F0 in the prosodic unit also increased accordingly, indicating that our ability in preplanning the pitch was closely related to the prosodic structure.

Chen and Xu [13] proposes that the neutral tones share a static tone target M, which we conjecture could be influenced by the position of the neutral tone in their reading materials. They put the neutral tone phrases at the end of prosodic phrase in the middle of carrier sentences. As a result, the boundary tone of the prosodic phrase might have an impact on the neutral tone realization. We suggest that the prosodic environment influences the F0 realization of the neutral tone target significantly (which will be addressed in another study).

Moreover, we found that with the increased number of neutral tone syllables, the initial F0 in a prosodic unit is also scaled higher accordingly, indicating that our ability in preplanning the pitch is closely related to prosodic structure.

5. ACKNOWLEDGEMENTS

This work is supported by the Key NSSFC (No. 15ZDB103) and CASS innovation project "Key Laboratory of Phonetics and Speech Science".

6. REFERENCES

[1] Chao, Y., *The Language Issues*, Beijing: The Commercial Press, 1968.

[2] Chao, Y., An intonational study of Beijing Mandarin, in *The Last Five Minutes*, appendix, Beijing: Zhonghua Book Company, 1929. Also published in Chao 2002, 253-272, 2002.

[3] Chao, Y., A preliminary study on the English intonation (with American variations) and the corresponding intonation in Chinese, *The Bulletin of Institute of History and Philology, Special Edition: Festschift in Honor of Prof. Yuanpei Cai on His 65th Birthday*, 1932. Also published in Chao 2002, 718-733, 2002.

[4] Lin, T., The relation between stress and syntactic structures in Modern Chinese, *Studies of the Chinese Language* 7, 301-334, 1962.

[5] Wang, J. and N. Smith, Studies in Chinese Phonology, Berlin: Mouton de Gruyter, 1997.

[6] Wang, J., Experimental phonetics, generative phonology, and studies on the pitch of neutral tone in Chinese, *Contemporary Linguistics* 4, 227-230, 2000.

[7] Lin, T., *Linguistic Essays By Lin Tao*, Beijing: The Commercial Press, 2001.

[8] Li, Z. A perceptual account of asymmetries in tonal alignment. In *Proceedings of the North East Linguistic Society* 33, eds. by Makoto Kadowaki and Shigeto Kawahara, Amherst, MA: UMass GLSA, 147-166, 2003.

[9] Chao, Y., Gwoyeu Romatzyh or the National Romanization, *Monthly on Mandarin* 7, 87-117, 1922. Also published in *Linguistic Essays By Yuenren Chao*, eds. by Wu. Z, Beijing: The Commercial Press, 61-72, 2002

[10] Chao, Y., Tone and intonation of Chinese, *The Bulletin of Institute of History and Philology* 4(3), 121-134, 1933. Also published in Chao 2002, 734-749, 2002.

[11] Chao, Y., *A Grammar of Spoken*, trans. by Lvy, S., Beijing: The Commercial Press, 1979. The original published by Berkeley, CA: California University Press, 1968.

[12] Luo, C. and Wang, Y., *An Introduction to General Linguistics*, Beijing: Science Press, 1957.

[13] Chen, Y. and Xu, Y., Production of weak elements in speech -- Evidence from f0 patterns of neutral tone in standard Chinese, *Phonetica* 63, 47-75, 2006.

[14] Lin, T., A preliminary study on the properties of neutral tone in Beijing Mandarin, *Essays on Linguistics* 10, Beijing:The Commercial Press, 16-37, 1983.

[15] Cao, J., The properties of neutral-syllables in Standard Chinese, *Journal of Applied Acoustics* 4, 3-8, 1986.

[16] Wang, Y., The roles of pitch and duration in the perception of neutral tone in Standard Chinese, *Chinese Journal of Acoustics* 29(5), 453-461, 2004.

[17] Li. A., Gao, J., Jia, Y., and Wang, Y., Pitch and duration as cues in perception of neutral tone under different contexts in Standard Chinese, *Proceedings of APSIPA*, Angkor Wat, 2014.

[18] Li, A. and Fan, S., Correlates of Chinese neutral tone perception in different contexts, *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK, 2015.

[19] Li, A. Phonetic Correlates of Neutral Tone in Different Information Structures, *Contemporary Linguistics* 19(3), 348-378, 2017.

[20] Sun, N., *F0 features of tone and boundary tone and their phonetic realization in Mandarin Chinese*, PhD thesis of CASS (in Chinese), 2006.

[21] Carlos, G., *The Phonology of Tone and Intonation*. Cambridge University Press, 2004.

[22] Chao, Y., *Linguistic Essays By Yuenren Chao*, Beijing: The Commercial Press, 2002.

[23] Xu, Y., Fundamental frequency peak delay in Mandarin, *Phonetica* 58, 26-52, 2001.

[This paper was published at O-COCOSDA 2018]