Prosodic realization of tonal target and F0 peak alignment in Mandarin neutral tone

Aijun Li and Zhiqiang Li
Chinese Academy of Social Sciences | University of San Francisco

Neutral tone in Mandarin is generally believed to lack tonal identity and exhibit more variability in its phonetic realization. We examined the tonal target of neutral tone in a prosodic word consisting of a full syllable (S) and one, two, or three neutral-tone syllables. In the experiment, the test words, presented in isolation and embedded in a carrier sentence, were read in two intonation patterns: declarative and interrogative. The results showed: (1) the tonal target of neutral tone is L(ow) at the end of the intonation phrase in declarative intonation and M(id) in question intonation; (2) its phonetic realization is influenced by intonation patterns, the tone of S and the number of neutral-tone syllables in the prosodic word; (3) the influence of the tone of S is more robust in shorter sequences than in longer ones with three neutral-tone syllables; (4) placement of the F0 peak in T2 (LH) and the neutral tone immediately following T3 (L) is susceptible to the number of neutral-tone syllables. It seems clear from our study that while the tonal target of neutral tone is related to prosodic structure, its actual F0 scaling is sensitive to prosodic manipulations such as intonation patterns and prosodic word length. In addition, *tonelessness* of neutral tone allows for more freedom in the alignment of the F0 peak, whose temporal coordination with its segmental host is, nevertheless, subject to both phonological and phonetic constraints.

Keywords: neutral tone, tonal target, peak alignment, intonation patterns, Mandarin Chinese

1. Introduction

Mandarin Chinese distinguishes full and weak syllables. A full syllable is stressed and carries one of the four lexical tones: Tone 1 (H), Tone 2 (LH), Tone 3 (L) and Tone 4 (HL). When a syllable is weak and unstressed, it does not carry a lexical tone, but is said in *neutral tone* (Chao 1968). For example, in disyllabic words like

https://doi.org/10.1075/lali.00099.li | Published online: 15 December 2021
Language and Linguistics 23:1 (2022), pp. 47–81. ISSN 1606-822X | E-ISSN 2309-5067

bōli 'glass', the first syllable is in Tone 1 and the second in neutral tone. A weak syllable is also called a neutral-tone or toneless syllable. It is generally perceived as short and lax.

Previous studies on neutral tone in Mandarin Chinese have approached it from a wide swath of perspectives, both phonetic and phonological, ranging from traditional auditory-based descriptions to experimental explorations (e.g. Dong 1958; Chao 1968; Lin & Yan 1980; Cao 1986; Lin 2012; Gao & Li 2018). Some examined its lexicological, syntactic, and phonological properties while others focused on acoustic realizations and perceptual correlates. The present study investigates the tonal target of neutral tone experimentally based on how it responds to prosodic manipulations by creating target words consisting of a full syllable and following neutral-tone syllables and setting them up in different prosodic contexts and intonation patterns. With this experimental setup, we are also hoping to explore in detail how the F0 peak – which is customarily mapped to H tone in phonological representations – is aligned with its syllabic host in the neutral-tone context.

We expect the results to bear on two broad issues in the phonology-phonetics interface. On the one hand, a central question is how abstract phonological structures are physically realized in real time and space. In this connection, a distinction is often made between phonological processes, which are dealt with in the phonological component of the grammar, and phonetic processes, which are dealt with in the phonetic implementation component. As it has been observed, mapping from the output of phonology to observable surface phonetic events is not always straightforward (e.g. Cohn 1990; Keating 1996). Rather phonetic processes exhibit both contextual and cross-linguistic variations. However, a growing body of research has shown that phonetic variations are not happening randomly; they are often constrained in different ways. One factor that has been found to regulate contextual variations is the perceptual distinctiveness of phonological contrasts (Hayes et al. 2004). The oft-cited example is Manuel (1990) in which she found that vowel-to-vowel coarticulation is more limited in more crowded vowel inventories, where neighboring vowel phonemes are less distinct acoustically.

On the other hand, a number of recent studies conclude that precise temporal coordination of F0 events with their segmental hosts is quite complex and that tonal alignment does not follow directly from phonological association. In phonology, the association of tone or pitch accent with specific elements of the segmental string (i.e. tone-bearing unit) is specified by autosegmental links: Chinese tones are associated with moraic segments (or syllable rimes), English pitch accents are associated with stressed syllables, and Japanese word accents are associated with a specific mora.

Such an association relation is often obscured in the phonetic implementation process where tonal targets are realized as F0 events. A cross-linguistically common pattern is that the F0 peak of a tone or pitch accent may appear after the tone-bearing unit (mora, rime, or syllable) with which it is phonologically associated. This phenomenon is sometimes dubbed *peak delay* (PD). For example, Silverman & Pierrehumbert (1990) found that in English the F0 peak in the prenuclear H* pitch accent often occurs after the accented syllable that bears the pitch accent. Prieto, van Santen & Hirschberg (1995) reported that in Mexican Spanish the F0 peak in H* accent is delayed unless an accented syllable follows. In their study of rising prenuclear accent in Modern Greek, Arvaniti, Ladd & Mennen (1998) found that the H target is consistently aligned just after the beginning of the first post-accentual vowel. Xu (2001) presented similar peak delay data for the Mandarin rising tone, Tone 2, in which the F0 peak mostly occurs to the right of the syllabic boundary and falls inside the following onset consonant in the context of a low tone such as Tone 3. The situation is further complicated by speech rate and the prosodic position of the tonal target in question (Steele 1986, Silverman & Pierrehumbert 1990). In addition, there seem to be consistent, but also fine-grained cross-linguistic variations in tonal target alignment. Ladd and his colleagues conducted a number of studies (Modern Greek: Arvaniti et al. 1998; British English: Ladd et al. 1999; Dutch: Ladd et al. 2000; German: Atterer & Ladd 2004), reporting consistent alignment patterns of tonal targets in the languages they studied; those languages exhibit small but significant alignment differences from one another. Specifically, although the low target is aligned with the onset of the stressed syllable in both British English and Modern Greek prenuclear rises, the high target is aligned earlier in the former case, typically late in the immediately following consonant. In German, the low target of the prenuclear rise is "aligned well within the initial consonant of the stressed syllable or even early in the stressed vowel", with Northern speakers aligning the low target earlier than Southern speakers (Atterer & Ladd 2004).

Neutral-tone syllables provide an intriguing context in which the F0 peak in the preceding lexical tone interacts with the tonal target (or lack thereof) of neutral tone, in that alignment of the F0 peak is sensitive to the following tonal context. The timing of the F0 peak also plays a role in the F0 scaling of the tonal target of neutral tone. It is exactly the interaction of tonal target and F0 peak alignment that guides the experimental design and motivates the analysis, to be presented in the remainder of this paper.

2. Basic properties of neutral tone: phonology and phonetics

As mentioned earlier, Mandarin Chinese distinguishes four lexical tones on stressed syllables. A stressed syllable carries one of the four lexical tones. In contrast, when a syllable is unstressed, it does not carry a lexical tone, and hence is in neutral tone or toneless. Some syllables, like suffixes and sentential particles, are always in neutral tone except in citation form when Tone 1 will be used instead. When stressed syllables become unstressed in certain lexical or prosodic contexts, they lose their underlying tones.

There are different types of neutral-tone syllables, depending on whether they are always toneless or derived via deletion of their lexically associated tones. A very small number of morphemes, such as perfective suffixes, question markers and sentence particles, are always unstressed and said in neutral tone. There is no way to identify their underlying tones. When said in isolation, they become stressed and take on Tone 1. Other neutral tone syllables originate from stressed syllables, but are said in neutral tone when used in certain grammatical or prosodic contexts. For example, the directional verbal ending *lai* is typically said in neutral tone, but it can be used as a verb in its stressed, toned form of Tone 2, meaning 'to come'. Some lexical categories, such as pronouns and prepositions, alternate between stressed-toned and unstressed-toneless forms, depending on the prosodic contexts in which they occur. Examples are given below. Most of them are taken from Dong (1958). Following standard pinyin convention (the PRC romanization system for Mandarin Chinese), tone diacritics appear above syllables (e.g. ā, á, ǎ, à representing the four tones respectively), with syllables in neutral tone left unmarked.

(1) a. suffixes: *zi, tou*, plurality marker *men*
examples: *yǐ zi* 'chair', *mù tou* 'wood', *wǒ men* 'we'

b. particles: possessive marker *de*, aspect marker (perfective) and sentence-final particle *le*, aspect marker (perfective) *guo*, aspect marker (progressive) *zhe*, modal particle *ba*
examples: *wǒ de* 'mine', *lái le* 'come+aspect marker', *hǎo ba* 'good+modal particle'

c. localizers: *li* 'inside', *shang* 'above'
examples: *wū li* 'inside the house', *tiān shang* 'in the sky'

d. pronouns as objects: *wo* 'I', *ni* 'you', *ta* 's/he'
examples: *zhǎo ni* 'look for you', *jiào wo* 'call me'

e. verbs reduplicated as cognate objects:
examples: *kàn kan* 'have a look', *shūo shuo* 'say it', *xiǎng xiang* 'think it over'

f. directional verbal endings: *lai* 'come', *qu* 'go out'
examples: *ná lai* 'bring here', *zǒu chu qu* 'walk out'

In a few disyllabic words, whether the second syllable is stressed or unstressed makes a lexical contrast. For these words, the metrical status of the second syllable is an idiosyncratic property of the word. In (2), each pair is distinguished by the second syllable: it is stressed in the first column and in neutral tone in the second.

(2) *mǎi mài* 'buying and selling' – *mǎi mai* 'business'
xíng lǐ 'salute' – *xíng li* 'luggage'
dōng xī 'east and west' – *dōng xi* 'thing'
bào chóu 'revenge' – *bào chou* 'payment'
láng tóu 'wolf head' – *láng tou* 'hammer'
dà yì 'outline' – *dà yi* 'careless'

In a few other disyllabic words, the second syllable is always unstressed and said in neutral tone, such as *pián yi* 'cheap', *páng xie* 'crab', *yá men* 'government office in feudal China' and *zhuó mo* 'to ponder'. Further, there are a small number of disyllabic words in which the second syllable can be optionally stressed or unstressed without incurring any semantic differences, such as *lǎo hǔ/hu* 'tiger' and *bō lí/li* 'glass'.

Neutral-tone syllables are unlike full-toned stressed syllables which can occur in any position within a word: they only occur in restricted positions: they are usually cliticized to the preceding stressed syllable. This distribution pattern entails that in a disyllabic word, the syllable in neutral tone must be the second one, forming a trochaic pattern.

It is generally assumed in traditional analyses that syllables in neutral tone do not have independent pitch values, and their phonetic pitch varies with the preceding lexical tone (e.g. Chao 1948, 1968; Dong 1958; Lin 1962, 1985). The phonetic pitch of neutral tone following four lexical tones is summarized below, transcribed in Chao's tone digits on a five-point scale (Chao 1930), on which 1 indicates the lowest pitch and 5 the highest pitch. Tone values of the four lexical tones are also provided. In most analyses, it is sufficient to distinguish two pitch levels in the description of neutral tone: high after Tone 3 and low after the other tones (Chao 1968: 36; Cheng 1973).

Table 1. Phonetic pitch of neutral tone

Context	Pitch	Example	Gloss
after Tone 1 (55)	2	*tā de*	'his'
after Tone 2 (35)	3	*huáng de*	'yellow one'
after Tone 3 (214)	4	*nǐ de*	'yours'
after Tone 4 (51)	1	*dà de*	'big one'

Acoustic properties of neutral-tone syllables in disyllabic words have been intensively studied in experimental reports (e.g. Lin & Yan 1980; Cao 1986; Wang 2004; Lu & Wang 2005; Lin 2012; Gao & Li 2018). It is found that on average, neutral-tone syllables are about half as long as the corresponding stressed syllables in the minimal pairs similar to those in (2). Perception experiments also reveal that duration plays a vital part in the correct identification of neutral tone syllables in adults and infants (Lin & Yan 1980; Lin 1985; Cao 1986; Li et al. 2014; Li & Fan 2015; Fan et al. 2018). The fundamental frequency (F0) patterns of neutral-tone syllables are found to be consistent with the above interpretation: mid after tone 3 and low after the other tones.

Studies of neutral tone in recent years have zeroed in on the acoustic and perceptual properties of neutral tone in varied prosodic contexts (Wang 1997, 2000; Li 2003a, 2003b; Wang 2004; Sun 2006; Chen & Xu 2006; Li 2017). Li Aijun (2017), for example, examined the realization of tonal target of neutral tone in disyllabic words in which the second syllable is in neutral tone or one of the four lexical tones in five different information structures. Figure 1 illustrates the F0 patterns of disyllabic words in which the first syllable is in one of the four tones (T1, T2, T3, T4) and the second syllable is in neutral tone (Tn).

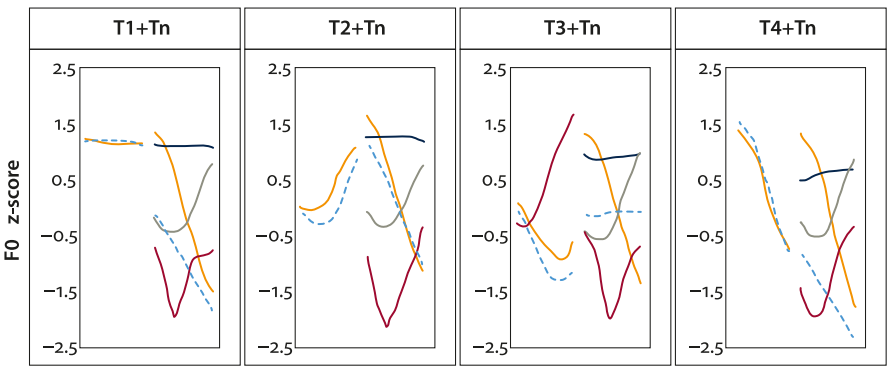


Figure 1. F0 contours of disyllabic words in which the second syllable is in one of the four tones (solid lines) vs. in neutral tone (dotted lines), grouped by the first syllable's tone. Each tone is normalized horizontally for tonal duration and z-scored normalized vertically for F0. (Li 2017)

In Figure 1, syllables in neutral tone are showing a mid-falling F0 contour after T1, a high-falling F0 contour after T2 and a low-falling F0 contour after T4. It is reasonable to assume that a low tone is aligned with the neutral-tone syllable and the falling F0 contour arises as transition from where the preceding full tone ends to the low tone target. After T3, a mid-level F0 contour appears instead. Several analyses have been proposed for the H tone (interpreted as such in the phono-

logical representation) on the neutral-tone syllable after a T3. Assuming T3 is L, Yip (1980) and Wang (1997) propose a rule of H insertion after L, while Duanmu (1999) considers the H on the neutral-tone syllable after a T3 originating from a polarity requirement that L should be followed by H in a disyllabic foot. Milliken (1989) suggests that T3 is L followed by a floating H. The floating H is realized on the neutral-tone syllable after a T3. For the purpose of the present study, the choice among the above analyses is not necessary. They all agree in one way or another that neutral tone itself takes the L following the other three tones. The presence of the L tone is blocked by the H insertion or re-association after a T3, resulting in the H on the neutral-tone syllable. In addition to analyses that consider it a tonal phenomenon, a recent acoustic study suggested that the higher pitch in the neutral-tone syllable after T3 is due to an articulation-based effect as a result of the preceding low pitch (Prom-on et al. 2012). It is worth pointing out that in the Tianjin dialect the syllable in neutral tone following T1 (low tone, similar to T3 in Mandarin Chinese) gets L pitch, as in *māma* 'mother', so the post-L raising seems to be due to a language- or dialect-specific effect.

Li Zhiqiang (2003a) presented a preliminary study of F0 timing when multiple neutral-tone syllables are involved in a constraint-based model of tonal target realization.

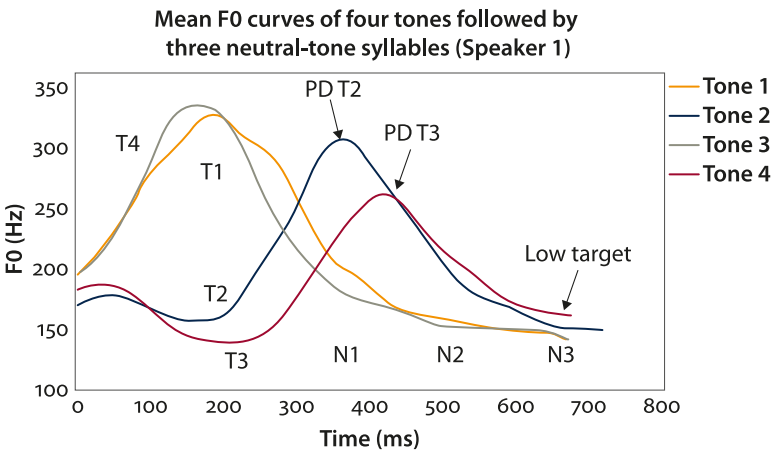


Figure 2. Mean F0 curves of four tones followed by three neutral-tone syllables (Li 2003a)

What is shown in Figure 2 is a prosodic word (PW) consisting of a full syllable in one of the four tones and three syllables in neutral tone, marked as N1, N2, and N3. An example is *māma men de* 'mother, plurality suffix, possessive suffix'. In terms of the tonal target, the F0 contours reach a clear low tone target at the end of the prosodic word. This key observation lends credence to the hypothesis that

neutral tone is toneless – the syllable does not carry one of the four lexical tones – and the low tone is associated with the end of a prosodic domain (Li 2003b). We call this the boundary L analysis of neutral tone to differentiate it from the default L analysis outlined above. Li (2003a) also discovered a significant amount of peak delay in the realization of T2 and T3 in a prosodic word with multiple neutral-tone syllables.

The study reported in Li (2003a) is partly motivated by an empirical difference between the F0 peak alignment in Modern Greek rising accent and Mandarin rising tone. As reported in Xu (2001), the F0 peak in Mandarin rising tone mostly occurs just inside the following onset segment. In Modern Greek, the F0 peak in a rising accent is aligned on average 15–20 ms after the beginning of the following unstressed vowel (Arvaniti et al. 1998). In addition, the onset of the rise is also aligned differently: in Modern Greek, it occurs right before the onset of the syllable, whereas in Mandarin, it occurs in the middle of the syllable rime. The difference could possibly arise from *tonal crowding* or *stress clash* in Mandarin because in Xu's study the rising tone was followed by a stressed syllable carrying a low tone (T3), while in Arvaniti et al.'s speech material there were always at least two unstressed syllables on either side of the test stressed syllable carrying the rising accent.

In yet another study, Chen & Xu (2006) propose that the neutral tone does have a specific underlying pitch target, which is likely to be static and mid, but it is likely to be implemented with a weak articulatory strength, as illustrated in Figure 3.

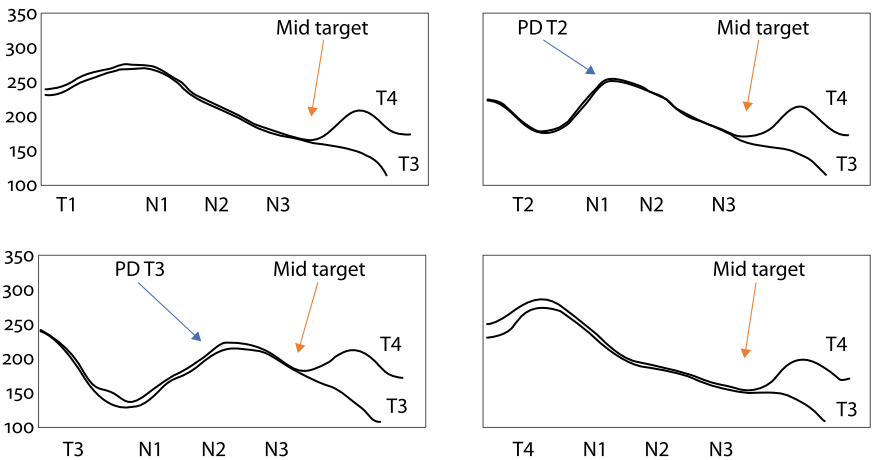


Figure 3. Neutral tone has a mid and static target tone in different tonal contexts. (Chen & Xu 2006)

Chen & Xu (2006) also report that T2 and T3 exhibit post-L F0 raising when followed by neutral-tone syllables. For example, after a T3, they found that the post-L F0 raising could reach the second neutral-tone syllable. In the test material used in their work, both tonal contexts before and after neutral-tone syllables were controlled for in the experimental design. Neutral-tone syllables were followed by two lexical different tones, T3 (L) and T4 (HL), as Figure 3 shows. While the target words in their study were only embedded in the middle of the sentence, our study examines the neutral tone in the sentence-final position, with a focus on the boundary tone effects contributed by statements and interrogative sentences and the F0 declination effect.

Our study seeks to answer the following questions:

1. What is the tonal target of neutral tone?
2. How is the tonal target realized in different tonal and prosodic contexts?
3. How is the F0 peak in the preceding lexical tone aligned with its syllabic host in the context of a varying number of neutral tones?
4. What are the contributing factors that constrain the phonetic realization of the neutral tone target?

3. Method

3.1 Experimental design and material

The test material is divided into two groups: words read in isolation and words embedded in a carrier sentence. See Table 2 for the complete list of target words, where S stands for a stressed full syllable and N for a neutral-tone syllable. The number after N refers to the position of neutral tone in the sequence.

The target words consist of a stressed syllable in one of the four lexical tones, followed by one, two, or three neutral-tone syllables. They are either kinship terms or constructed out of kinship terms. For example, the tonal sequence in *mèimei* 'little sister' is T4N1. The plural form is *mèimeimen* 'little sisters', after attaching the plurality suffix *men*, resulting in the tonal sequence of T4N1N2. The possessive form is derived by adding the possessive suffix *de* – *mèimeimende* 'little sisters'' – and its tonal sequence is T4N1N2N3. The two suffixes are always said in neutral tone. For the purpose of our study, we consider all words in group A prosodic words, which will be read in isolation. Group B includes all words in group A and words that end with the sentence-final particle (PAR) *le*, which is also always said in neutral tone. Morphologically, words like *mèimei le* 'little sister PAR' are neither noun or noun phrase, but they form a prosodic word when embedded in a carrier sentence.

Table 2. Target words read in isolation (A) and in a carrier sentence (B)

Context	Tones	SN1	SN1N2	SN1N2N3
A:	T1	*māma* 'mother'	*māmamen* 'mothers'	*māmamende* 'mothers' '
			māmade 'mother's'	
	T2	*yéye* 'grandpa'	*yéyemen* 'granpas'	*yéyemende* 'grandpas' '
			yéyede 'grandpa's'	
	T3	*nǎinai* 'grandma'	*nǎinaimen* 'grandmas'	*nǎinaimende* 'grandmas' '
			nǎinaide 'grandma's'	
	T4	*mèimei* 'little sister'	*mèimeimen* 'little sisters'	*mèimeimende* 'little sisters' '
			mèimeide 'little sister's'	
B:	T1	*māma* *mā le* 'mother PAR'	*māmamen*	*māmamende* *māmamen le* 'mothers PAR'
			māmade	
			māma le 'mother PAR'	
	T2	*yéye*	*yéyemen*	
			yéyede	*yéyemende*
		yé le 'grandpa PAR'	*yéye le* 'grandpa PAR'	*yéyemen le* 'grandpas PAR'
	T3	*nǎinai*	*nǎinaimen*	
			nǎinaide	*nǎinaimende*
		nǎi le 'grandma PAR'	*nǎinai le* 'grandma PAR'	*nǎinaimen le* 'grandmas PAR'
	T4	*mèimei*	*mèimeimen*	
			mèimeide	*mèimeimende*
		mèi le 'little sister PAR'	*mèimei le* 'little sister PAR'	*mèimeimen le* 'little sisters PAR'

We used the same carrier sentence as in Sun (2006), which literally means "that bowl of flour was sent to ___". Two examples are given in (3) in which the test words *mā le* and *māmamen le* are underlined. The prosodic structure of the carrier sentence is provided, too.

(3) $[_{IP}\ [_{PP1}$ nà wǎn miànfěn$]\ [_{PP2}\ [_{PW}$ sòng gěi$]\ [_{PW}$ SN1$]]]$
Nà wǎn miànfěn sòng gěi <u>mā le / māmamen le</u>
That bowl flour send give mother PAR / mothers PAR
'That bowl of flower was sent to mother/mothers.'

3.2 Participants

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

3.3 Recording procedure

The target words in the two groups – with and without a carrier sentence – were divided into three groups randomly; each was read by five male and five female speakers. All words were read in two intonation patterns, declarative intonation and question intonation. The recording was conducted at a sampling rate of 16KHz with a 16bit quantitative resolution in a sound-proof booth in the phonetics lab at the Institute of Linguistics affiliated with the Chinese Academy of Social Sciences. The recordings obtained were monitored and screened by a lab assistant to make sure they were of good quality and without mispronunciations. In the end, there were 440 sentences read in declarative intonation and 560 in question intonation, 1,000 sentences in total, among which 440 were short sentences with the target words read in isolation and 560 were long sentences with the target words read in the same carrier sentence.

3.4 Data analysis

All sound files were first automatically segmented by an alignment program to generate syllable-level and phone-level boundaries. Visual inspection and manual correction were taken to ensure that the segmental boundaries were accurately marked. Prosodic information such as prosodic structure and stress was manually labeled. All annotations were created by two professional transcribers in the lab. An annotated utterance is given in Figure 4. The annotation includes *segmental boundaries*, syllables (the Pinyin tier), initials and finals (the IF tier), and break indexes (the BI tier) labelling prosodic word, prosodic phrase, and intonation phrase; and stress indexes (the ST tier) labelling prosodic word stress, prosodic phrase stress, and sentence-level stress.

F0 contours of each recording were analyzed by Praat and manually checked for spurious cycles. F0 values were then extracted for each final in a syllable at ten points in equal intervals to normalize in duration dimension, using a Praat script, with creaky voice excluded.

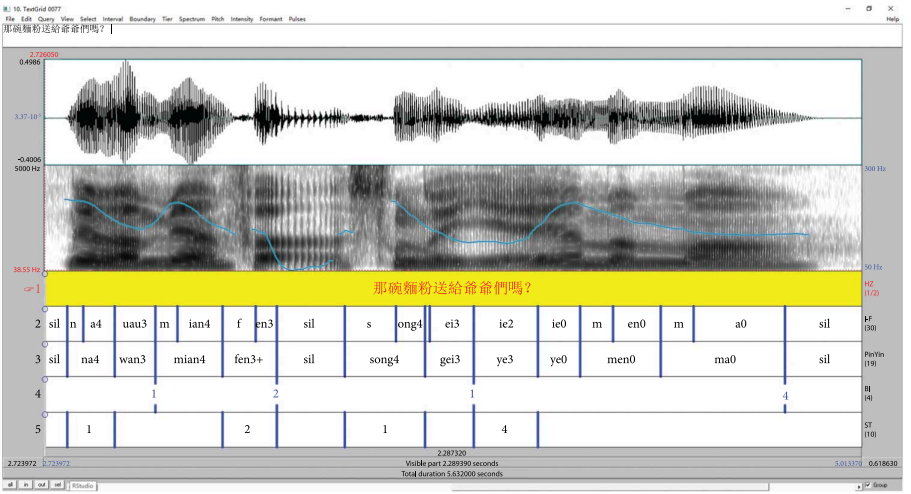


Figure 4. An example of data annotation

In order to eliminate speakers' individual differences, the F0 data of each speaker were normalized into a Z-score scale value (z_{lf0}), using the formula below.

$$Z_{lf0_i} = \frac{If0_i - m_{lf0}}{S_{lf0}} = \frac{\log_{10} f0_i - \frac{1}{n}\sum_{j=1}^{n}\log_{10} f0_j}{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(\log_{10} f0_i \frac{1}{n}\sum_{k=1}^{n}\log_{10} f0_k)^2}} \quad (4)$$

Where z_{lf0i} is the normalized z-score the i^{th} F0 value, m_{lf0} the mean value and s_{lf0} the standard deviation of the logarithmic F0 values respectively. n is the total number of F0 value of the individual speaker.

4. Results and analysis

4.1 F0 patterns and tonal target in declarative intonation

We first present the mean and normalized F0 contours of the target words read in isolation (a) and in a carrier sentence (b) in Figures 5–7. These words are read in declarative intonation. In each figure (Figures 5–10), the tonal duration is normalized by extracting eleven points for each syllabic final in equal steps, where the occurrence of F0 peak delay in T2 and T3 is also marked.

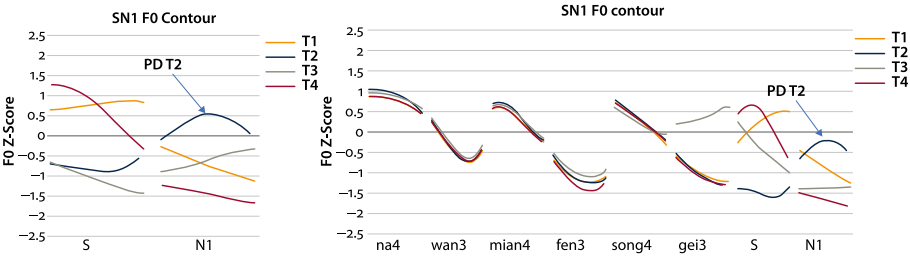
Figure 5a demonstrates that in the SN1 sequences read without a carrier sentence, the F0 contour on the neutral-tone syllable is dependent on the F0 of the preceding tone. The finding is consistent with traditional descriptions and more

recent experimental study (e.g. Li 2017). The ending F0 values in N1, do not seem to converge toward a specific point, but they tend to point to a low tone target after T1, T2, and T4. The F0 peak in T2 is realized in the following N1, well into the middle of the syllable; whereas the F0 peak in T1 stays within its syllabic boundary. In other words, there is no peak delay in T1. This is likely to be the reason why the ending F0 in N1 after T2 is much higher than that after T1: given the late occurrence of the F0 peak in T2, the falling F0 contour does not have enough time to reach a lower F0 (cf. Sundberg 1979). The H tone is realized in the neutral-tone syllable after T3, as expected from the phonological analyses laid out earlier. Similar F0 patterns are observed in the SN1 sequences read in a carrier sentence. For example, the separation of ending F0 values after T1 and T2 is also clearly observed in the neutral-tone syllable at the end of the sentence when peak delay happens in T2.

In the SN1N2 sequences, as shown in Figure 6a–b, the ending F0 values all converge to a low tone target in N2 after T1, T2, and T4. The F0 contour in N1 does not seem to correspond to clearly identifiable tone targets and they are more susceptible to the influence of the tonal contexts on both sides. For example, in the T1N1N2 sequence, the falling F0 contour in N1 starts from the F0 offset in T1 and ends at the F0 onset in N2, which is realized as a low tone target. A similar pattern is seen in the T4N1N2 sequence too. In the T2N1N2 sequence, however, the F0 peak in T2 is again realized in the following N1 as in the T2N1 sequence. The difference is that when there are two neutral-tone syllables following T2, the F0 contour is gradually declining from the F0 peak in N1 all the way to the end of N2, reaching the low tone target. In the T3N1N2 sequence, the F0 peak associated with the H tone in the neutral tone following T3 shows up in N2: well into the N2 territory in the isolation reading and at the beginning of N2 in the carrier-sentence reading. As a result, the F0 contour in N1 becomes transitional, lacking a clear tonal target. Another consequence of the late peak in the T3N1N2 sequence is that the F0 contour in N2 does not have much time to reach the low tone target at the end, especially given that neutral-tone syllables have much reduced duration. The prediction here is that in the T3N1N2N3 sequence, the low tone target will be reached N3 when the F0 peak is aligned with N1. This is exactly what happens in Figure 7.

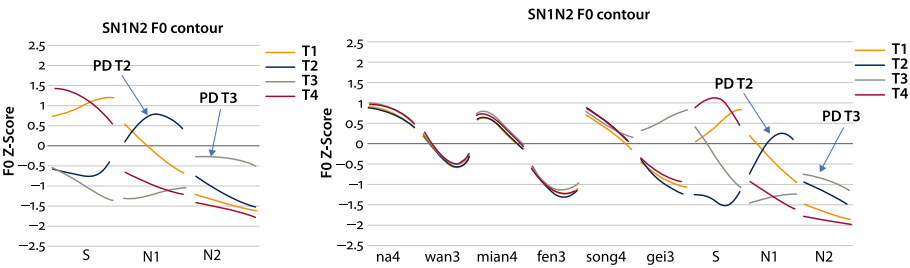
As the number of neutral-tone syllables increases in the SN1N2N3 sequences, the final neutral-tone syllable N3 is approximating to a canonical low tone target, aligned with the end of the sequence, as Figure 7a–b shows. As in the SN1N2 sequences, the F0 contours on N1N2N3 after T1 and T4 are gradually declining to the final low target, while the F0 peak delay happens in the sequences headed by T2 and T3. Similar patterns are observed in the target words read in a carrier sentence with the higher F0 offset in the final neutral-tone syllable following T3

than the other three tones, as seen in Figure 7b. Given the position of the F0 peak following T3 and the shorter duration available for the falling F0 to fully manifest itself, it is not surprising to witness this kind of separation, which is also noticeable in the SN1N2 sequences in Figure 6.



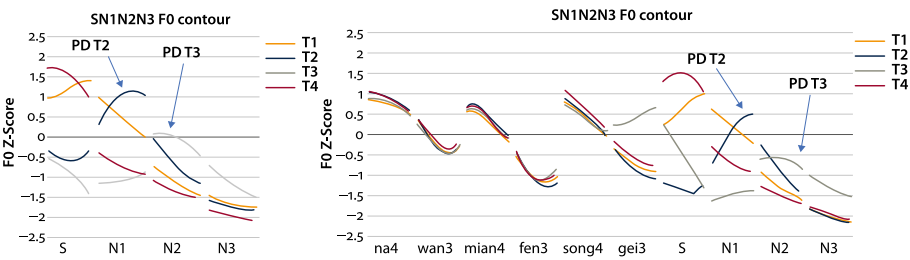
a. **b.**

Figure 5. Mean and normalized F0 contours of SN1 read in isolation (a) and in a carrier sentence (b) in declarative intonation



a. **b.**

Figure 6. Mean and normalized F0 contours of SN1N2 read in isolation (a) and in a carrier sentence (b) in declarative intonation



a. **b.**

Figure 7. Mean and normalized F0 contours of SN1N2N3 read in isolation (a) and in a carrier sentence (b) in declarative intonation

In summary, we have found that in multiple neutral-tone sequences such as SN1N2N3, a low tone target is aligned with the final neutral-tone syllable. The F0 contours in the neutral-tone syllables in-between do not seem to correspond to any clearly identifiable tonal targets. Rather, they are interpolated from where the preceding full tones end to the final low tone target. Its actual F0 scaling correlates with duration – from the F0 peak to the low target at the end of the prosodic word – available to fully realize the falling F0. In the case of T4N1, the low target is easily reached at the end of the neutral-tone syllable because T4 in Mandarin Chinese is a high falling tone that has an early F0 peak, from where the falling F0 contour continues all the way to the end of the sequence. This pattern of F0 contours is quite strong in all sequences headed by T4. Since the F0 peak in T1 occurs late in the syllable, the falling F0 contour does not quite reach the low target in T1N1 as in the case of the T4 sequences. However, when more duration becomes available, as in T1N1N2 and T1N1N2N3, the F0 converges to the lowest point of the pitch range, reaching the low target.

The F0 scaling of the low target in the sequences headed by T2 and T3 is influenced by the peak delay, i.e. the placement of the F0 peak. In the case of T2, the F0 peak invariably surfaces on the immediately following neutral-tone syllable N1, regardless of the number of neutral-tone syllables in the sequence. In the case of T3, the F0 peak was often realized on the next neutral-tone syllable in the T3N1 sequence, but it could occur on the N2 or even N3 as in the T3N1N2 and T3N1N2N3 sequences. When the placement of the F0 peak is too close to the end of the prosodic word as in SN1 or SN1N2, there is not enough time for the F0 contour to fully reach the low tone target aligned at the end.

4.2 F0 patterns and tonal target in question intonation

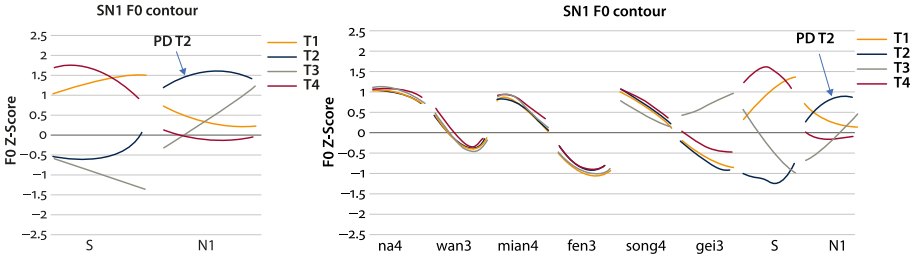
In this section, we present data obtained when the target words were read in question intonation, first in isolation and then in a carrier sentence. The mean and normalized F0 contours of the target words read in isolation (a) and in a carrier sentence (b) are shown in Figures 9–11. The occurrence of F0 peak delay in T2 and T3 is also marked. It should be pointed out that in Mandarin Chinese, a yes-no question can be formed by attaching the question morpheme *ma* or *ba*, both in neutral tone, to a declarative sentence. The questions created this way are typically said in declarative intonation. To elicit the question intonation, we asked the subject to read the target words as an echo question. For example, the example in (3) can be uttered as a question by simply adding a question mark at the end of the sentence. The implied meaning is, are you sure about that?

(5) *Nà wǎn miànfěn song gěi mā le / māmamen le?*
That bowl flour send give mother PAR / mothers PAR?
'Was that bowl of flower sent to mother/mothers?'

It has been suggested that a key difference between declarative intonation and question intonation in Mandarin Chinese lies in the lowering or raising effect of the F0 contour at the end of an intonation phrase due to the effect of boundary tone (e.g. Lin 2012). For example, the F0 contour of T2 (LH) at the end of an intonation phrase is further elevated, a more fully implemented rising tone, in question intonation, but realized with a compressed pitch range in declarative intonation.

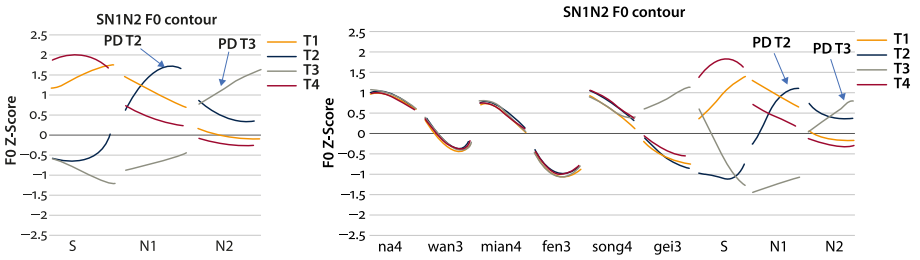
Several observations can be made here. First, as the number of neutral-tone syllables increases, the F0 contours in the final neutral-tone syllables tend to converge to a mid-tone target, especially in the SN1N2N3 sequences. This happens when the target word was read in isolation or in a carrier sentence. Second, the peak delay in T2 and T3 is also attested in the question intonation. In general, the F0 peak in T2 is aligned mostly with N1 and occasionally N2 while that in T3 often occurs in N2 and N3. Third, the F0 contours in the neutral-tone syllables are determined by the preceding full tone and the mid-tone target in the final neutral-tone syllable. For example, in the T1N1N2 sequence, the F0 peak in T1 appears at the end of the stressed syllable, from which point the F0 contour starts to decline gradually to mid-target. Similarly, in the T4N1N2N3 sequence, the F0 contour starts the declining trajectory from the F0 peak in T4 all the way to mid-target aligned with N3. Lastly, the scaling of the mid-tone target is also influenced by the peak delay in T2 and T3. In the T2N1 sequence, the F0 peak is delayed into N1, followed by a much truncated falling F0 toward the mid-tone target. In the T3N1N2 sequence, the F0 peak is reached at the end of N2, thus blocking the mid-tone target. In the T3N1N2N3 sequence, the F0 contour reaches its peak in N2 and then moves toward the mid-tone target.

By comparing the F0 patterns in declarative intonation and question intonation, we propose that a low tone target is aligned with the neutral-tone syllable at the end of the prosodic word, which is at the right edge of the intonational phrase in our experimental setup. The pitch height of the low tone is determined by several factors, including preceding tones, placement of the F0 peak, the number of neutral-tone syllables and the intonation patterns. When uttered in question intonation, the pitch height of the low tone is elevated to the mid tone.



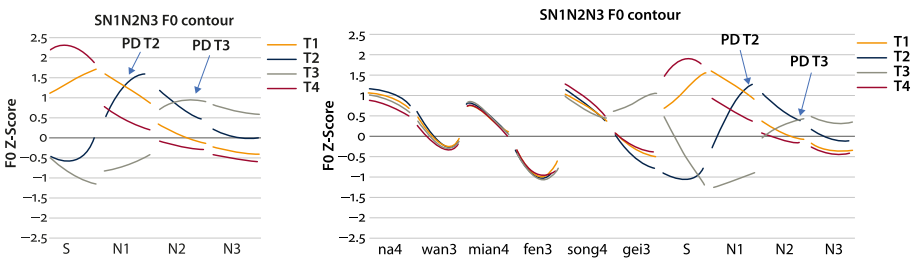
a. **b.**

Figure 8. Mean and normalized F0 contours of SN1 read in isolation (a) and in a carrier sentence (b) in question intonation



a. **b.**

Figure 9. Mean and normalized F0 contours of SN1N2 read in isolation (a) and in a carrier sentence (b) in question intonation



a. **b.**

Figure 10. Mean and normalized F0 contours of SN1N2N3 read in isolation (a) and in a carrier sentence (b) in question intonation

4.3 Statistical analyses of the tonal target in neutral tone

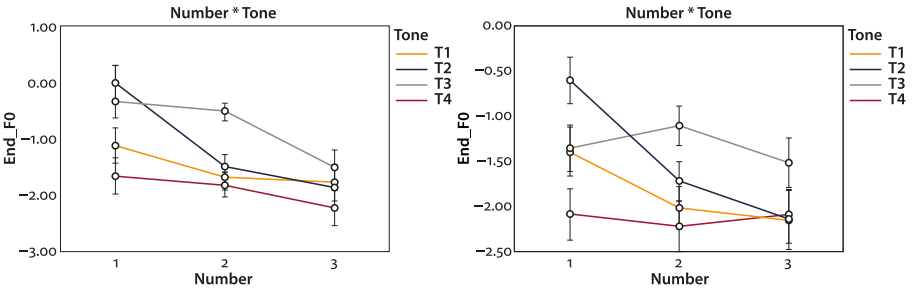
In order to examine the interactive impact of the prosodic contexts on the tonal target of the domain-final neutral-tone syllable (i.e. the F0 target value in the last neutral tone), three statistical methods were adopted. General Linear Mixed Model (GLMM) best suits the data in the current study (fixed effects factors are the number of neutral tones and the initial full tones; random effects factor is the sentence or speaker ID). One-way ANOVA was conducted to examine specific within-group differences in different numbers of neutral-tone syllables. Repeated measures design was used in analyzing the SN1N2N3 sequences. In what follows, we present the results obtained for the target words tested in all four conditions: the target words are read in isolation and then in a carrier sentence in the declarative intonation and in question intonation, respectively.

4.3.1 *LMM analysis*

LMM analysis results are showed in Figure 11–12 for the target words read in isolation and in a carrier sentence, in declarative intonation, and in question intonation, respectively. In declarative intonation, the number of neutral-tone syllables (Number), the initial lexical tone in the sequence (Tone) and their Interaction are significant in predicting the tonal target F0 scaling of neutral-tone syllables both in isolation (Number, $F(2,138)=48.157$, $p<0.001$; Tone, $F(3,138)=33.882$, $p<0.001$; Interaction, $F(6,138)=8.871$, $p<0.001$) and in a carrier sentence (Number, $F(2,201)=32.568$, $p<0.001$; Tone, $F(3,201)=17.206$, $p<0.001$; Interaction, $F(6,201)=9.263$, $p<0.001$).

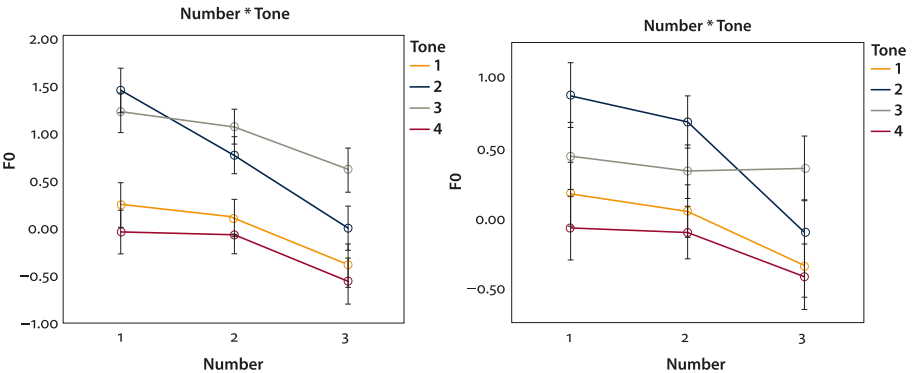
In question intonation, Number, Tone and their Interaction are also found to be significant both in isolation (Number, $F(2,267)=46.968$, $p<0.001$; Tone, $F(3,267)=103.336$, $p<0.001$; Interaction, $F(6,267)=9.648$, $p<0.001$) and in a carrier sentence (Number, $F(2,268)=19.563$, $p<0.001$; Tone, $F(3,268)=37.88$, $p<0.001$; Interaction, $F(6,268)=4.313$, $p<0.001$).

Based on the analysis, the tonal identity of the initial full syllable (S) in the sequence and the number of neutral-tone syllables both have a significant impact on the F0 scaling of the tonal target in the final neutral-tone syllable. The pitch height in the neutral-tone syllable immediately after T3 is higher than those after the other three tones. The tonal target in the final neutral-tone syllable in declarative intonation is much lower than that in question intonation.



a. **b.**

Figure 11. Interaction between the number of neutral-tone syllables and the initial tones for the tonal target F0 scaling in isolation (a) and in a carrier sentence (b) in declarative intonation



a. **b.**

Figure 12. Interaction between the number of neutral-tone syllables and the initial tones for the tonal target F0 scaling in isolation (a) and in a carrier sentence (b) in question intonation

4.3.2 *ANOVA analysis*

One-way ANOVA was carried out to further examine within-group differences in the F0 scaling of the tonal target in the last neutral-tone syllable in order to determine the effect of the initial lexical tones and the number of neutral-tone syllables in the target words in the four test conditions.

In declarative intonation, when the target words were read in isolation, we found the following. Please refer to Figure 11a for the visual presentation of the results. (1) In the SN1 sequences, the F0 target values of N1 show significant differences following the four lexical tones ($F(3, 35) = 16.162$, $p < 0.001$). Pairwise,

significant differences exist in the sequences headed by T1 and T2 ($p = 0.01$), T1 and T3 ($p = 0.026$), T4 and T2 ($p < 0.01$), and T4 and T3 ($p < 0.01$). No significant difference is found in the T1 and T4 sequences, and T2 and T3 sequences, respectively. In other words, T2 and T3 pattern together and stand in contrast to T1 and T4 in terms of their impact of the F0 scaling of N1. It is worth nothing that the F0 peak in T1 and T4 materialize within the syllabic boundary of their segmental hosts, much earlier than the F0 peak in T2, which is often delayed into N1, and the F0 peak after T3, which is always realized in N1. (2) In the SN1N2 sequences, the F0 target values of N2 show significant differences following the four lexical tones ($F(3,71) = 43.983$, $p < 0.001$). Specifically, N2 following T3 shows significantly higher F0 than those following T1, T2 and T4, and the difference is significant at the level of $p < 0.01$. No significant difference is found among T1, T2, and T4, though. (3) In the SN1N2N3 sequences, no significant difference is found in the F0 target values of N3 following all four tones ($F(2,32) = 2.515$, $p = 0.076$). What this result entails is that in the isolation reading condition in declarative intonation, the final neutral tone is gradually approaching a stable low tone target, as the number of neutral-tone syllables is increasing. By the same token, the influence on the tonal target of neutral tone from the initial full tones in the target words is also diminishing.

When the target words were read in a carrier sentence in which the target words form the last prosodic unit, we have identified generally similar statistical tendencies, with results visually presented in Figure 11b. (1) In the SN1 sequences, the F0 target values of N1 show significant differences following the four lexical tones ($F(3,63) = 15.871$, $p < 0.001$). Pairwise, significant differences exist in the sequences headed by T1 and T2 ($p = 0.02$), T1 and T4 ($p = 0.028$), T2 and T3 ($p < 0.01$), and T2 and T4 ($p < 0.01$). The F0 scaling of the neutral tone in descending order is: T2>T1, T3>T4. (2) In the SN1N2 sequences, the F0 target values of N2 also show significant differences following the four lexical tones ($F(3,88) = 22.230$, $p < 0.001$). The same pattern emerges in which the F0 in N2 scaling significantly higher in the T3N1N2 sequences than T1, T2 and T4, which do not show significant difference among themselves. (3) In the SN1N2N3 sequences, the result shows a significant difference in tones ($F(3,51) = 8.552$, $p < 0.001$), in which N3 following T3 is significantly higher than those following T1, T2 and T4. Specifically, significant difference is found in pairwise comparisons between T3 and the other three tones: T1 and T3 ($p = 0.01$), T2 and T3 ($p = 0.01$) and T4 and T3 ($p = 0.03$). Upon closer look, it has become clear that in the three neutral-tone sequences, the F0 trajectory is approaching the low tone target, which manifests almost identical F0 following T1, T2 and T4, as shown in Figure 11b. In the case of T3, the F0 of N3 is higher after T3 than after other three tones. A similar trend is also observed in the isolation reading condition,

discussed above, but the difference does not reach a statistically significant level. Here in the carrier sentence condition, the difference is significant, but only at the $p < 0.05$ level.

Which begs the question: why is T3 special? As we discussed in § 4.1, T3 is different from the other three lexical tones in terms of placement of the F0 peak. We have argued that the F0 peak after T3 is phonologically associated with the immediately following neutral-tone syllable, N1, while the F0 peaks in T1, T2 and T4 are either realized within their syllabic boundary (T1 and T4) or delayed into the following syllable (T2). In the SN1N2N3 sequences, the F0 peak after T3 occurs in N2 or even in N3. Further discussion of this issue will be offered in § 4.4.

The results are presented visually in Figure 12 for question intonation in which the target words were read in isolation (a) and in a carrier sentence (b). We discuss sequences in varying lengths separately. When the target words were read in isolation, we found the following. (1) In the SN1 sequences, the F0 target values of N1 show significant differences following the four lexical tones ($F(3,76) = 33.373$, $p < 0.001$), with N1 showing clearly higher F0 after T2 and T3 than after T1 and T4. (2) In the SN1N2 sequences, the F0 target values of N2 also show significant differences following the four lexical tones ($F(3,116) = 29.524$, $p < 0.001$). The grouping effect is also clear with N2 showing clearly higher F0 after T2 and T3 than after T1 and T4, but the caveat is that the F0 values of N2 following T2 and T3 show significant difference ($p < 0.05$). (3) In the SN1N2N3 sequences, the F0 values of N3 are showing significant differences following the four lexical tones ($F(3,75) = 26.023$, $p < 0.001$). While N3 following T3 shows significantly higher F0 than those following T1, T2, and T4 at the statistically significant level ($p < 0.001$), N3 following T2 is moving toward the lower F0 values in the direction of N3 after T1 and T4.

When the target words were read in a carrier sentence, they form the last prosodic unit. The results, visually presented in Figure 12b, paint a similar picture as in the isolation reading condition. (1) In the SN1 sequences, the F0 in N1 shows divergent values following the four tones ($F(3,75) = 14.093$, $p < 0.001$), in the descending order of T2>T3, T1>T4, but statistically significant differences only exist in pairwise comparisons of T4 and T2 ($p < 0.001$), T4 and T3 ($p < 0.05$), T1 and T2 ($p < 0.001$), and T3 and T2 ($p < 0.05$). (2) In the SN1N2 sequences, the F0 in N2 shows significant differences following the four tones ($F(3,114) = 12.857$, $p < 0.001$), in the descending order of T2>T3>T1, T4. Pairwise, the differences between T2 and T3 ($p < 0.05$), T2 and T1 ($p < 0.001$), and T2 and T4 ($p < 0.001$) have reached statistical significance. One interesting fact to note in the SN1N2 sequences is the higher F0 of N2 after T2. A quick visual inspection of the middle series in Figure 11–12 reveals that the F0 of N2 is higher after T3 in the two conditions in declarative intonation and in the isolation reading condition in ques-

tion intonation. This is because a much higher percentage (30%) of the F0 peak in T2 in the SN1N2 sequences is delayed into N2 when the target words are read in a carrier sentence in question intonation, as seen in Table 4, B. The late peak in N2 prevents the F0 from going down to a lower value in such a short duration in a neutral-tone syllable. (3) In the SN1N2N3 sequences, the F0 values of N3 are showing significant differences following the four lexical tones ($F(3,76) = 8.585$, $p < 0.001$). Like the isolation reading condition, N3 following T3 shows significantly higher F0 than those following T1, T2, and T4 at the statistically significant level ($p < 0.001$), N3 following T2 is moving toward the lower F0 values in the direction of N3 after T1 and T4.

4.3.3 *Repeated measurement*

The tonal target of the final neutral-tone syllables in the SN1N2N3 sequences is worth a closer look in order to determine the effect of the preceding lexical tones on the F0 scaling of N3 across different conditions. Specifically, we classified the target words in two Tone groups: those headed by T3 (T3 group) and those headed by other tones (non-T3 groups). A 2*2 (isolation/carrier sentence * tone groups) design was employed.

In declarative intonation, the repeated measurement shows that the F0 scaling of N3 did not significantly differ in the isolation reading and carrier sentence conditions ($F(1,105) = 0.034$, $p = 0.855$), while Tone groups did show significant difference ($F(1,105) = 12.854$, $p = 0.001$), with the T3 group significantly higher than the non-T3 groups. Their interaction is not significant ($F(1,105) = 0.009$, $p = 0.924$)), and as a result, t-test was not performed.

In question intonation, the repeated measurement shows that the F0 scaling of N3 did not significantly differ in the isolation reading and carrier sentence conditions either ($F(1,155) = 1.35$, $p = 0.247$), while Tone groups did show significant difference ($F(1,155) = 69.255$, $p < 0.001$), with T3 group significantly higher than the non-T3 groups. Their interaction is not significant ($F(1,155) = 2.236$, $p = 0.137$)), therefore, t-test was not performed.

The effect of intonation is confirmed in pairwise comparisons given in Table 3 for the T3 group and non-T3 groups read in isolation and in a carrier sentence, respectively. Recall that we argue that the tonal target for neutral tone is a low tone aligned with the end of the prosodic word in declarative intonation, and it is elevated to a mid tone in question intonation due to the effect of boundary tone when the prosodic word resides at the end of the intonational phrase. The side-by-side comparisons showed that in all conditions we examined, the F0 target values in the domain-final neutral-tone syllables are significantly different in declarative intonation and question intonation, at the level of $p < 0.001$ across the board. To wit, they are scaled much higher in question intonation. The sta-

tistical analysis has confirmed our analysis that the tonal target of neutral tone in the multiple neutral-tone context is realized consistently as the lowest F0 in the prosodic domain, thus carrying a low tone target in declarative intonation, which is subsequently elevated to mid tone in question intonation.

Table 3. Pairwise comparisons of tonal targets of neutral tone in T3 group and non-T3 groups in the three neutral tone sequences, A: read in isolation; B: read in the carrier sentence

		In declarative intonation		In question intonation			
	Tone of S	Mean	Sd	Mean	Sd	T-value	Sig. level
A:	T1/T2/T4	−1.90	0.42	−0.32	0.49	−14.266	$p<0.001$
	T3	−1.51	0.55	0.60	0.5	−10.559	$p<0.001$
B:	T1/T2/T4	−1.93	0.504	−0.23	0.51	−16.209	$p<0.001$
	T3	−1.52	0.49	0.35	0.6	−10.422	$p<0.001$

4.4 Placement of the F0 peak

We analyzed the placement of the F0 peaks in T2 and T3 and the frequency of peak delay in different prosodic contexts: in isolation, in a carrier sentence, in declarative intonation, and in question intonation. Recall that F0 peak delay refers to the phenomenon that an F0 peak sometimes occurs after the syllable with which it is associated either lexically or prosodically (Xu 2001), as schematically illustrated in Figure 13. In our study, peak delay is found in both T2 and T3. It happens in T2 (LH) when the F0 peak in the rising tone occurs in the following syllable. Strictly speaking, the H tone in the neutral-tone syllable immediately after a T3 is phonologically associated with that syllable, hence no peak delay. Peak delay occurs when the F0 peak is realized after N1 as in the T3N1N2 sequence.

The data on the peak delay in T2 and T3 when the sequences were read in declarative intonation, in isolation (A) and in the carrier sentence (B), are summarized in Table 4. In the T2 sequences in both A and B conditions, peak delay happens with its F0 peak falling almost exclusively in N1. Similarly, in the T3 sequences, the F0 peak occurs almost exclusively in N2, with only one exception in the sequence of T3N1N2 in B, where it stays in N1. As the number of neutral-tone syllable increases in longer sequences, especially in the SN1N2N3 sequences, the F0 peak can even encroach on N2 (30% in B in the case of T2) and N3 (15% in B in the case of T3).

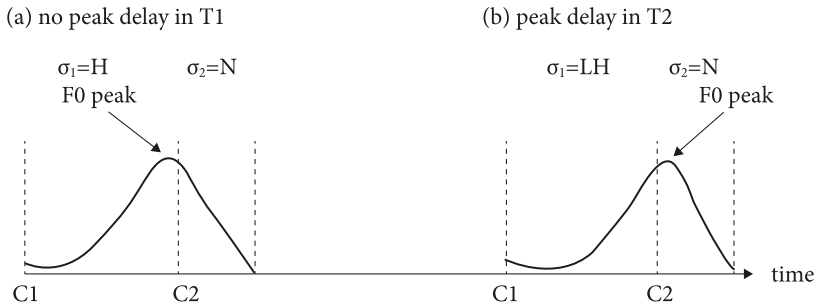


Figure 13. Schematic illustration of peak delay (Li 2003b), C1 marks the beginning of the syllable in T1 or T2, and C2 marks the beginning of the syllable in neutral tone.

Table 4. Placement of the F0 peaks in T2 and T3 in declarative intonation, A: read in isolation; B: read in the carrier sentence

		T2	N1	N2	N3	T3	N1	N2	N3
A:	SN1		10 (100%)				10 (100%)		
	SN1N2		20 (100%)					20 (100%)	
	SN1N2N3		9 (90%)	1 (10%)				9 (90%)	1 (10%)
B:	SN1		20 (100%)				20 (100%)		
	SN1N2		29 (97%)	1 (3%)			1 (3%)	29 (97%)	
	SN1N2N3		14 (70%)	6 (30%)				17 (85%)	3 (15%)

In question intonation, there seem to be more extensive peak delay in both T2 and T3, especially in longer sequences, as the data in Table 5 show. For example, the F0 peak falls regularly in N3 of the three neutral-tone words headed by T3: 45% in A and 75% in B. The F0 peak in T2 is realized in N2 more frequently in B: 30% in T2N1N2 and 55% in T2N1N2N3.

Table 5. Placement of the F0 peaks in T2 and T3 in question intonation, A: read in isolation; B: read in the carrier sentence

		T2	N1	N2	N3	T3	N1	N2	N3
A:	SN1		20 (100%)				20 (100%)		
	SN1N2		30 (100%)				1 (3%)	29 (97%)	
	SN1N2N3		18 (90%)	2 (10%)				11 (55%)	9 (45%)
B:	SN1		20 (100%)				20 (100%)		
	SN1N2		21 (70%)	9 (30%)				30 (100%)	
	SN1N2N3		9 (45%)	11 (55%)				5 (25%)	15 (75%)

4.5 Durational analyses of F0 peak alignment

Three durational measurements were taken in order to determine exactly how the F0 peaks in T2 and T3 are aligned with their segmental hosts in different prosodic contexts. The relevant points of measurement are illustrated in Figure 14.

(6) a. duration ratio (DR) of peak delay (pd) in the neutral-tone syllable (syl)
DR_syl = (T_peak – T_onset_Ni) / Dur_Ni (i = 1, 2, 3)
b. duration of peak delay (Dur_pd)
Dur_pd = T_peak – T_onset_S
c. duration ratio (DR) of peak delay (pd) in the whole word (word)
DR_word = Dur_pd / Dur_word

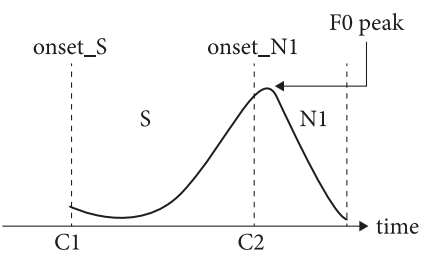


Figure 14. Measurements of duration

Table 6. Duration ratio of peak delay in the neutral-tone syllable after T2 and T3 in declarative intonation, A: read in isolation; B: read in the carrier sentence

DR_syl (mean/std)	T2	N1	N2	N3	T3	N1	N2	N3
A:	SN1		0.374/ 0.144			0.728/ 0.222		
	SN1N2		0.557/ 0.175				0.384/ 0.197	
	SN1N2N3		0.612/ 0.219	0.164/ 0			0.517/ 0.171	0.066/ 0
B:	SN1		0.423/ 0.173			0.523/ 0.284		
	SN1N2		0.650/ 0.16	0.128/ 0		0.776/ –	0.428/ 0.25	
	SN1N2N3		0.720/ 0.251	0.142/ 0.075			0.517/ 0.184	0.333/ 0.286

Duration ratio of peak delay in the neutral-tone syllable measures how far the F0 peaks in T2 and T3 encroach into the following neutral-tone syllables, relative to the duration of neutral-tone syllable, in which the F0 peak is located. A ratio of 0.5 in N1 means the F0 peak is aligned with the mid point in the syllable. The higher the ratio, the further the peak is delayed into the following syllable. The data obtained in declarative intonation and in question intonation are presented in Tables 6 and 7 respectively, again under two conditions, read in isolation (A) and in a carrier sentence (B). One clear trend that emerges in both T2 and T3 sequences is that the alignment of the F0 peak can delay further into the neutral-tone syllable after the syllable with which it is phonologically associated, when the target words become longer with more neutral-tone syllables. In the meantime, more delay is found in condition B, i.e. when the target words are read in a carrier sentence, than in condition A. However, this trend is more obvious in declarative intonation than in question intonation. In general, question intonation tends to create a context for later F0 peaks, especially in the sequences headed by T3. In Table 6, the F0 peak after T3 appears in the second half of N2, 0.82 and 0.83 in the T3N1N2 sequences, 0.68 and 0.86 in the T3N1N2N3 sequences in conditions A and B. When the F0 peak is realized beyond the immediately following neutral-tone syllable, i.e. in N2 in the case of T2 and in N3 in the case of T3, it always stays within the first half of the syllable duration.

Table 7. Duration ratio of peak delay in the neutral-tone syllable after T2 and T3 in question intonation, A: read in isolation; B: read in the carrier sentence

DR_syl (mean/std)		T2	N1	N2	N3	T3	N1	N2	N3
A:	SN1		0.541/ 0.280				0.960/ 0.064		
	SN1N2		0.741/ 0.142				0.758/ –	0.822/ 0.118	
	SN1N2N3		0.716/ 0.187	0.325/0.03				0.676/ 0.143	0.308/ 0.285
B:	SN1		0.595/ 0.273				0.930/ 0.160		
	SN1N2		0.792/ 0.161	0.167/ 0.091				0.834/ 0.129	
	SN1N2N3		0.694/ 0.271	0.261/ 0.128				0.855/ 0.119	0.360/ 0.335

Duration of peak delay measures displacement in seconds of the F0 peak from the onset of the syllable it is phonologically associated with. Although the H tone is phonologically associated with the neutral-tone syllable after T3 in line with phonological analyses, the measurement was taken from the onset of T3 in order to make comparisons with T2. Like before, we present the results of the target words read in isolation and in a carrier sentence, in declarative intonation and in question intonation.

In declarative intonation (Figure 15), the alignment of the F0 peak in T2 is quite stable in the sense that when it lands in N1, the whole rising F0 contour lasts on average 290 ms regardless of the number of neutral-tone syllables. When the peak appears in N2, the average duration of the whole rising F0 contour is 335 ms. The alignment of the F0 peak in T3 is more sensitive to the number of neutral-tone syllables, with the average duration of the whole rising F0 contour at 460 ms when the peak occurs in N2 and at 570 ms when the peak occurs in N3. There is no peak delay when the F0 peak is realized in N1 after T3.

In question intonation (Figure 16), the average duration of the whole rising F0 contour is becoming increasingly shorter as the number of neutral-tone syllables increases in the T2 sequences, ranging from 268 ms to 350 ms. Although the variation is small, we do not have a good explanation for the trend. When the peak is realized in N2, the average duration of the rise is 335 ms. The alignment of F0 peak in T3 tends to be more sensitive to the number of neutral-tone syllables. The

average duration of the rise is in the range of 500 ms to 540 ms when the peak is in N2. When the peak is in N3, the whole rise extends its duration to 580 ms.

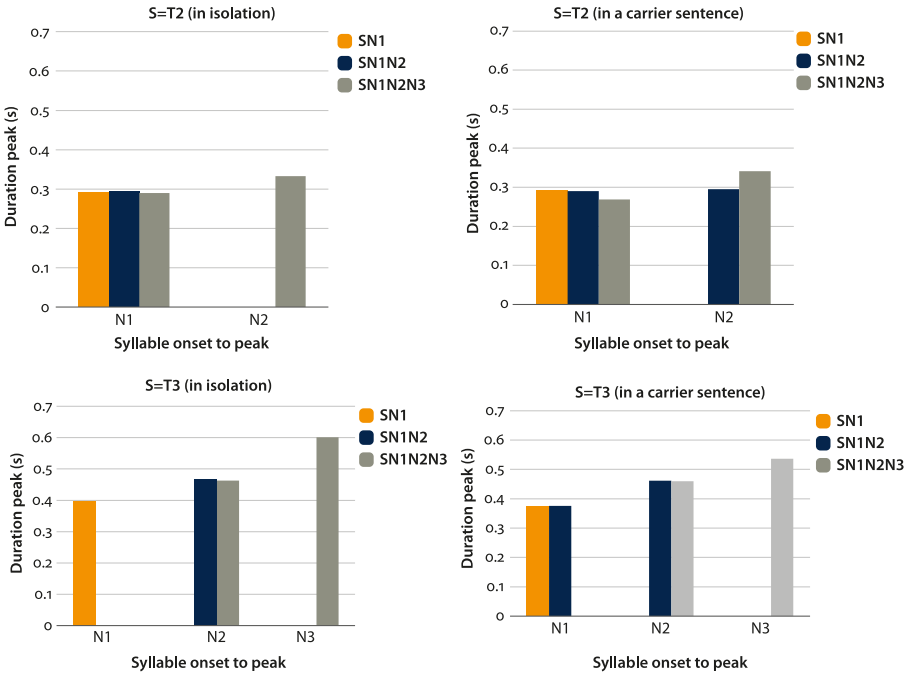


Figure 15. Alignment of the F0 peaks in T2 and T3, measured from syllable onset, in declarative intonation

Duration ratio of peak delay in the whole prosodic word measures the placement of the F0 peak relative to the duration of the whole word – the SN1, SN1N2 and SN1N2N3 sequences – headed by T2 and T3. It is an indicator of where the F0 peak is aligned in the whole word. The results are summarized in Table 7 and Table 8 respectively, for the test words read in declarative and question intonation.

Examination of the data does not seem to reveal a strong correlation between the placement of F0 peak and the number of neutral-tone syllables, or equivalently, the length of the target word. While in declarative intonation (Table 8), the duration ratio is bigger in condition B than in condition A, the trend is not as robust in question intonation (Table 9). It is safe to conclude, at least based on the data in our work, that the whole prosodic word is not a reliable indicator as the specific neutral-tone syllable (N1, N2, or N3) with regard to F0 timing.

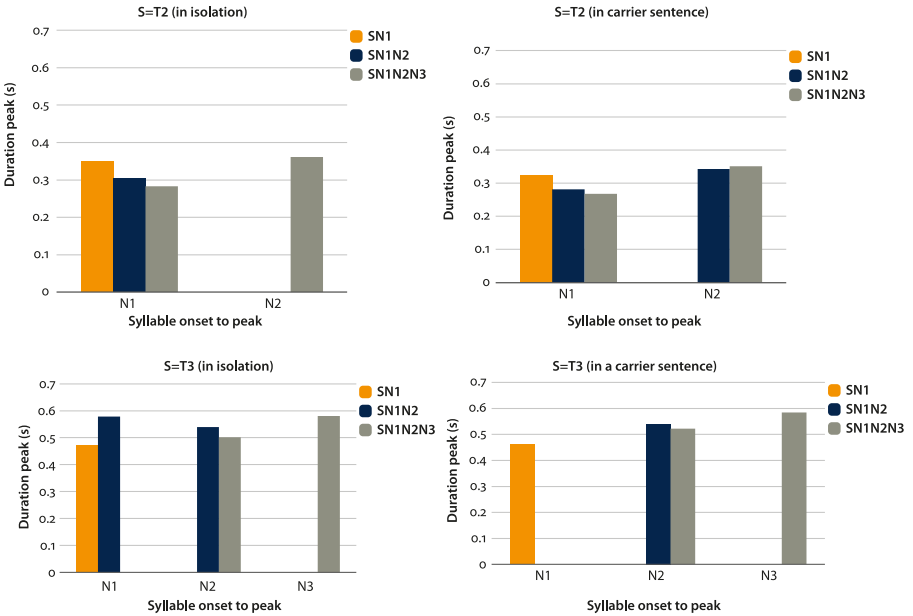


Figure 16. Alignment of the F0 peaks in T2 and T3, measured from syllable onset, in question intonation

Table 8. Duration ratio of peak delay in the whole word headed by T2 and T3 in declarative intonation, A: read in isolation; B: read in the carrier sentence

DR_word (mean/std)		T2	N1	N2	N3	T3	N1	N2	N3
A:	SN1		0.708/ 0.081				0.860/ 0.118		
	SN1N2		0.496/ 0.072					0.786/ 0.084	
	SN1N2N3		0.396/ 0.038	0.487/ 0				0.614/ 0.063	0.809/ 0
B:	SN1		0.747/ 0.085				0.795/ 0.130		
	SN1N2		0.559/ 0.079	0.697/ 0			0.624/ –	0.813/ 0.093	
	SN1N2N3		0.408/ 0.044	0.510/ 0.032				0.633/ 0.067	0.827/ 0.050

Table 9. Duration ratio of peak delay in the whole word headed by T2 and T3 in question intonation, A: read in isolation; B: read in the carrier sentence

DR_word (mean/std)		T2	N1	N2	N3	T3	N1	N2	N3
A:	SN1		0.764/ 0.154				0.943/ 0.156		
	SN1N2		0.503/ 0.048				0.910/ –	0.925/ 0.050	
	SN1N2N3		0.367/ 0.047	0.485/ 0.003				0.601/ 0.064	0.819/ 0.098
B:	SN1		0.780/ 0.152				0.966/ 0.077		
	SN1N2		0.499/ 0.052	0.634/ 0.043				0.933/ 0.059	
	SN1N2N3		0.380/ 0.053	0.477/ 0.060				0.624/ 0.076	0.826/ 0.103

5. Conclusion

This study investigates the tonal target of neutral tone experimentally in different tonal and prosodic contexts. The four variables that we have focused on are the number of neutral-tone syllables (SN1 vs. SN1N2 vs. SN1N2N3), lexical tones of the preceding full syllable, two intonation patterns (declarative and interrogative), and two reading conditions (in isolation and in a carrier sentence). The empirical questions we set out to explore are the following:

1. What is the tonal target of neutral tone?
2. How is the tonal target realized in different tonal and prosodic contexts?
3. How is the F0 peak in the preceding lexical tone aligned with its syllabic host in the context of a varying number of neutral tones?
4. What are the contributing factors that constrain the phonetic realization of the tonal target of neutral tone?

Our analysis concludes that there is a clearly identifiable tonal target in the neutral-tone syllable and it is L(ow). This conclusion should be qualified in several ways. First, it is not the case that every neutral-tone syllable receives an L tone in the phonological representation or post-lexical process. The prosodic words used in this study contain a full syllable (S) and neutral-tone syllables (N) from

one to three. In the SN1, SN1N2, and SN1N2N3 sequences, the L is aligned with the end of the prosodic word. To wit, only the PW-final neutral-tone syllable receives the L tone. The F0 contours in the non-final neutral-tone syllables is determined jointly by the preceding tone and the domain-final L tone. The phonetic mechanism to generate transitional F0 is F0 interpolation, which can be linear or asymptotic. Second, the L tone is realized often as a non-canonical L due to prosodic factors such as intonation, but its alignment is clearly with the end of the syllable in question.

The phonetic realization of the domain-final L tone is constrained by both phonology and phonetics. The L emerges as M when the prosodic word is said in question intonation. This is not to say that we are dealing with two tones. Phonologically, the neutral tone at the end of the prosodic word is aligned with L, which is realized as M at the end of the intonation phrase in question intonation due to the effect of boundary tone. The F0 scaling of the L is constrained in other ways too. The location of the F0 peak in the preceding tone and the length of the prosodic word are both relevant. In shorter sequences like SN1 or SN1N2, the F0 contour does not have enough duration to reach the L tone, especially when the F0 peak is delayed, but in the longer strings like SN1N2N3, the F0 almost invariably converges to the L tone at the end of the prosodic word. In our study, the length of a prosodic word varies with the number of neutral-tone syllables.

Neutral-tone syllables provide a tonal context which facilitates peak delay. The F0 peak in T2 routinely appears in N1 or even N2 in longer sequences. Compared with the T2-T3 setup used in Xu (2001) to study peak delay in T2, more peak delay is happening when T2 is followed by neutral-tone syllables. In the T2N1N2 and T2N1N2N3 sequences, N1 in the former and N1N2 in the latter are not associated with any tones. Their respective "tonelessness" allows a little more space for the F0 peak in T2 to encroach further into the following syllables thanks to its articulatory dynamics. In the T2-T3 context, the L tone in T3 would prevent the F0 peak from going too much into the following syllable in T3. A crucial point to make here is preservation of tonal contrast (Li 2003a). A late F0 peak into the syllable rime would create a high falling F0 contour in the syllable, which could potentially cause confusion with T4. Here is an example in which constraints on tonal contrast operate to restrict the placement of F0 targets.

Peak delay also happens when T3 is followed by two or three neutral-tone syllables. The F0 peak appears in the neutral-tone syllable immediately following T3 due to the H tone either in the phonological representation of T3 or created in the phonological process. L in T3 and H in the immediately following neutral-tone syllable create a rising F0. In the T3N1N2 and T3N1N2N3 sequences, the F0 peak mostly occurs in N2 and N3. The number of neutral-tone syllables is a strong indicator in predicting where the F0 peak is aligned.

Neutral tone presents an interesting case study which allows us to take a closer look at the mapping from phonology to phonetics. It seems reasonable to assume that phonological and phonetic constraints play an important role in the mapping process, which, in our case, manifests itself in how the tonal target of neutral tone is defined and realized as a function of the prosodic structure, and how it interacts with the F0 peak alignment within the prosodic word.

Abbreviations

ANOVA	analysis of variance	PAR	sentence-final particle
BI	break index	PD, pd	peak delay
DR	duration ratio	PRC	the People's Republic of China
F0	fundamental frequency	PW	prosodic word
H	high	S	full syllable
IF	initial and final	Spk	speaker
L	low	ST	stress index
LMM	linear mixed model	syl	syllable
M	mid	T	tone
N	neutral tone		

References

Arvaniti, Amalia & Ladd, D. Robert & Mennen, Ineke. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26(1). 3–25. https://doi.org/10.1006/jpho.1997.0063

Atterer, Michaela & Ladd, D. Robert. 2004. On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics* 32(2). 177–197. https://doi.org/10.1016/S0095-4470(03)00039-1

Cao, Jianfen. 1986. Putonghua qingsheng yinjie texing fenxi [An analysis of properties of neutral tone syllables in Standard Chinese]. *Yingyong Shengxue* [Journal of Applied Acoustics] 5(4). 1–6.

Chao, Yuen Ren. 1930. A system of tone letters. *Le Maître Phonétique* 45(30). 24–27.

Chao, Yuen Ren. 1948. *Mandarin primer: An intensive course in spoken Chinese*. Cambridge: Harvard University Press. https://doi.org/10.4159/harvard.9780674732889

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Yiya & Xu, Yi. 2006. Production of weak elements in speech – Evidence from F0 patterns of neutral tone in standard Chinese. *Phonetica* 63. 47–75. https://doi.org/10.1159/000091406

Cheng, Chin-chuan. 1973. *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton. https://doi.org/10.1515/9783110866407

Cohn, Abigail. 1990. *Phonetic and phonological rules of nasalization*. Los Angeles: University of California. (Doctoral dissertation.)

Dong, Shaowen. (ed.). 1958. *Yuyin changshi* [Common sense in phonetics]. Beijing: Culture and Education Press.

Duanmu, San. 1999. Metrical structure and tone: Evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics* 8(1) 1–38. https://doi.org/10.1023/A:1008353028173

Fan, Shanshan & Li, Aijun & Chen, Ao. 2018. Perception of lexical neutral tone among adults and infants. *Frontiers in Psychology* 9. (https://www.frontiersin.org/article/10.3389/fpsyg.2018.00322) (Accessed 2019-01-20.) https://doi.org/10.3389/fpsyg.2018.00322

Gao, Jun & Li, Aijun. 2018. Production of neutral tone on disyllabic words by two-year-old Mandarin-speaking children. In Fang, Qiang & Dang, Jianwu & Perrier, Pascal & Wei, Jianguo & Wang, Longbiao & Yan, Nan. (eds.), *Studies on speech production* (Lecture Notes in Computer Science 10733), 89–98. Cham: Springer. https://doi.org/10.1007/978-3-030-00126-1_9

Hayes, Bruce & Kirchner, Rovert & Steriade, Donca. (eds.). 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486401

Keating, Patricia. 1996. The phonology-phonetics interface. *UCLA Working Papers in Phonetics* 92. 45–60.

Ladd, D. Robert & Faulkner, Dan & Faulkner, Hanneke & Schepman, Astrid. 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106(3). 1543–1554. https://doi.org/10.1121/1.427151

Ladd, D. Robert & Mennen, Ineke & Schepman, Astrid. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107(5). 2685–2696. https://doi.org/10.1121/1.428654

Li, Aijun. 2017. Putonghua butong xinxijiegou zhong qingsheng de yuyin texing [Phonetic correlates of neutral tone in different information structures]. *Dangdai Yuyanxue* [Contemporary Linguistics] 19(3). 348–378.

Li, Aijun & Gao, Jun & Jia, Yuan & Wang, Yaru. 2014. Pitch and duration as cues in perception of neutral tone under different contexts in Standard Chinese. In Asia-Pacific Signal and Information Processing Association (ed.), *Proceedings of 2014 APSIPA Annual Summit and Conference, Siem Reap, city of Angkor Wat, Cambodia, December 9–12, 2014.* (http://www.apsipa.org/proceedings_2014/Data/paper/1037.pdf) (Accessed 2019-01-20.) https://doi.org/10.1109/APSIPA.2014.7041529

Li, Aijun & Fan, Shanshan. 2015. Correlates of Chinese neutral tone perception in different contexts. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. (https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0201.pdf) (Accessed 2019-01-20.)

Li, Zhiqiang. 2003a. A perceptual account of asymmetries in tonal alignment. In Kadowaki, Makoto & Kawahara Shigeto (eds.), *Proceedings of the North East Linguistic Society 33: Massachusetts Institute of Technology*, 147–166. Amherst: GLSA, University of Massachusetts Amherst.

Li, Zhiqiang. 2003b. *The phonetics and phonology of tone mapping in a constraint-based approach*. Cambridge: MIT. (Doctoral dissertation.)

Lin, Tao. 1962. Xiandai haiyu qingyin he jufa jiegou de guanxi [The relation between neutral tone and syntactic structure in Modern Chinese]. *Zhongguo Yuwen* [Studies of the Chinese Language] (7). 301–334.

Lin, Tao. 1985. Tantao Beijinghua qingyin xingzhi de chubu shiyan [Preliminary experiments in the exploration of the nature of Mandarin neutral tone]. In Lin, Tao & Wang, Lijia (eds.), *Beijing yuyin shiyanlu* [Working papers in experimental phonetics], 1–26. Beijing: Peking University Press.

Lin, Maocan & Yan, Jingzhu. 1980. Beijinghua qingsheng de shengxue xingzhi [Acoustic characteristics of neutral tone in Beijing Mandarin]. *Fangyan* [Dialect] 1980(3). 166–178.

Lin, Maocan. 2012. Hanyu yudiao shiyan yanjiu [The experimental study of intonation in Mandarin Chinese]. Beijing: China Social Sciences Press.

Lu, Jilun & Wang, Jialing. 2005. Guanyu qingsheng de jieding [On defining "qingsheng"]. *Dangdai Yuyanxue* [Contemporary Linguistics] 2005(2). 107–112.

Manuel, Sharon Y. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88(3). 1286–1298. https://doi.org/10.1121/1.399705

Milliken, Stuart. 1989. Why there is no third tone sandhi rule in Standard Mandarin. (Paper presented at the Tianjin International Conference on Phonetics and Phonology, Tianjin, 7–10 June 1989.)

Prieto, Pilar & van Santen, Jan & Hirschberg, Julia. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23(4). 429–451. https://doi.org/10.1006/jpho.1995.0032

Prom-on, Santitham & Liu, Fang & Xu, Yi. 2012. Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling. *Journal of the Acoustical Society of America* 132(1). 421–432. https://doi.org/10.1121/1.4725762

Silverman, Kim & Pierrehumbert, Janet. 1990. The timing of prenuclear high accents in English. In Kingston, John & Beckman, Mary E. (eds.), *Papers in the laboratory phonology I: Between the grammar and the physics of speech*, 72–106. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.005

Steele, Shirley A. 1986. Nuclear accent F0 peak location: Effects of rate, vowel, and number of following syllables. *Journal of the Acoustical Society of America* 80 (Supplement 1). S51. https://doi.org/10.1121/1.2023842

Sundberg, Johan. 1979. Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7(2). 71–79. https://doi.org/10.1016/S0095-4470(19)31040-X

Sun, Nianhao. 2006. Putonghua shengdiao yu bianjiediao yingao tezheng jiqi shixian guizhe [F0 features of tone and boundary tone and their phonetic realization in Mandarin Chinese]. Beijing: Chinese Academy of Social Sciences. (Doctoral dissertation.)

Wang, Jialing. 1997. The representation of the neutral tone in Chinese Putonghua. In Wang, Jialing & Smith, Norval (eds.), *Studies in Chinese phonology*, 157–183. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110822014

Wang, Jialing. 2000. Shiyan yuyinxue shengcheng yinxixue yu hanyu qingsheng yingao de yanjiu [Experimental phonetics, generative phonology and the study of the pitch of neutral tone in Chinese]. *Dangdai Yuyanxue* [Contemporary Linguistics] 2000(4). 227–230.

Wang, Yunjia. 2004. Yingao he shichang zai Putonghua qingsheng zhijue zhong de zuoyong [The effects of pitch and duration on the perception of the neutral tone in standard Chinese]. *Shengxue Xuebao* [Chinese Journal of Acoustics] 2004(5). 453–461.

Xu, Yi. 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* 58(1–2). 26–52. https://doi.org/10.1159/000028487

Yip, Moira Jean. 1980. *The tonal phonology of Chinese*. Cambridge: MIT. (Doctoral dissertation.)

Authors' addresses

Aijun Li (corresponding author)
Institute of Linguistics
Chinese Academy of Social Sciences
No. 5 Jianguomennei dajie
Beijing 100732
China

liaj@cass.org.cn

Publication history

Date received: 11 March 2019
Date accepted: 5 May 2020
Published online: 15 December 2021