Appeared in *Studies in the Linguistic Sciences* 26.1/2: 95-115

Pre-juncture Lengthening and Foot Binarity*

San Duanmu
University of Michigan

Abstract

This paper studies the lengthening of stressed pre-juncture syllables in English and Mandarin Chinese. First, I report a phonetic experiment which shows that pre-juncture lengthening is present in both languages but it is much greater in English than in Chinese. Then, I suggest that the result is expected if all full Mandarin syllables have the uniform structure CVX (Duanmu 1993).

1. Introduction

Pre-juncture (or pre-boundary) lengthening refers to the phenomenon where the final syllable before a major syntactic juncture is longer than a similar one in other positions. The phenomenon has also been called 'phrase-final lengthening' and 'pre-pausal lengthening'. Phrase-final lengthening may suggest that all XP final positions have a similar lengthening effect, which is not the case. Pre-pausal lengthening may suggest that the lengthening takes place only before a pause, which is not the case either. For example, Umeda (1975) found that in read English speech 'pre-pausal' lengthening was more frequently followed by no pause than followed by a pause. To avoid such ambiguities, I use the term 'pre-juncture lengthening' in this article.

Pre-juncture lengthening has been well documented by a large body of phonetic work in many languages. Some studies have also been made on comparing pre-juncture lengthening across languages. For example, Delattre (1965) suggested that English, French, German, and Spanish differed in their durational patterns at word junctures. Berkovits (1991) suggested that utterance-final lengthening was more evident in English than in Hebrew. Ao & Shih (1994) suggest that in Chinese there is phrase-final lengthening but not utterance-final lengthening.

In this study I compare pre-juncture lengthening in Chinese and English, with an emphasis on the phonological perspectives. Specifically, I am interested in the following questions: Is there a phonological basis for pre-juncture lengthening? In what ways is a difference in pre-juncture lengthening between two languages related to their phonological systems? Given a difference in pre-juncture lengthening, what other phonological differences can one expect?

For concreteness, I examine the lengthening of a stressed pre-juncture syllable in American English (hereafter English) and Mandarin Chinese (hereafter Chinese). In section 2 I present some phonetic results that English is more prone to pre-juncture lengthening than Chinese, in agreement with earlier findings (Umeda 1975, Streeter 1978, Shen 1992, Ao & Shih 1994). In section 3, I offer an explanation for the lengthening difference. Following Burzio (1994), I assume that a stressed foot should be (at least) two syllables long. Since both English and Chinese have left-headed feet, a stressed final syllable needs to be followed by another syllable. In English, the second syllable can be realized by lengthening the stressed final syllable, as a result of which it serves as two syllables. In Chinese, however, such lengthening is not available, and the second syllable is realized as a silent one (pause). The lengthening difference between English and Chinese is due to a difference in their syllable inventories. English has flexible syllable structures, such as CVC, CVV, V, VC, etc., whereby lengthening does not create an ill-formed syllable. For example, the monosyllabic [fæn] can turn into disyllabic [fæ.æn], [fææ.n], or [fææ.æn] without violating syllabic constraints. In contrast, Chinese has a fixed structure CVX for all regular syllables, so that [fæn] cannot be lengthened to [fæ.æn], [fææ.n], or [fææ.æn]; in the first both syllables are ill-formed and in the second and third the second syllable is ill-formed. Concluding remarks and further issues are given in section 4.

2. Phonetic Experiment

It has been noted in the literature that there is considerably difference in pre-juncture lengthening between English and Chinese. For example, Umeda (1975) found that in read English speech a vowel is about 70%-100% longer in a pre-juncture position than in other positions, whereas Ao & Shih (1994) found that in read Chinese speech a syllable is only about 20% longer in a pre-juncture position. Similarly, Streeter (1978) found that, in structures like '(A plus E) times O' vs. 'A plus (E times O)', English speakers can indicate syntactic boundaries by pre-juncture lengthening without pauses. In contrast, Shen (1992) found that in Chinese, although some degree of lengthening was present before a syntactic boundary, an additional pause was also present, and it was the pause that served as the major cue for syntactic boundaries.

While previous studies showed consistent results, they were based on fairly small numbers of speakers. For example, Umeda (1975) was based on three English speakers, Ao & Shih (1994) was based on one Chinese speaker, Streeter (1978) was based on one English speaker, and Shen (1992) was based on two Chinese speakers. In addition, these studies were primarily designed either for English or for Chinese, and not for comparing them directly. It will be helpful, therefore, to see whether the same result obtains with more speakers and under direct comparisons. For this purpose an experiment was conducted to compare pre-juncture lengthening in English and Chinese in controlled environments.

2.1. Method

2.1.1. Test sentences

Three sentences were used in English, shown in (1), with phonetic transcription. Three corresponding sentences were used in Chinese, shown in (2), with phonetic transcription, characters, tone, Pinyin transcription, word-for-word gloss, and translation. The tonal markings are: H = high (the First tone), R = rise (the Second tone), L = low (the Half-Third tone), F = fall (the Fourth tone), and 0 = toneless.

(1) a. an1 an2 an3
[tɛn plʌs tɛn taimz tɛn]
'(Ten plus ten) times ten.'

b. bn1 bn2 bn3
[tɛn plʌs tɛn taimz tɛn]
'Ten plus (ten times ten).'

c. [hi bɔt ə fæn bʌt ai bɔt ə kar]
'He bought a fan, but I bought a car.'

(2) a.

	an1		an2		an3
	[sæn	tɕa	sæn	ʈʂʰəŋ	sæn]
	(三	加	三)	乘	三
	H	H	H	R	H
	san	jia	san	cheng	san
	three	plus	three	time	three

'(Three plus three) times three'.

b.

	bn1		bn2		bn3
	[sæn	tɕa	sæn	ʈʂʰəŋ	sæn]
	三	加	(三	乘	三)
	H	H	H	R	H
	san	jia	san	cheng	san
	three	plus	three	time	three

'Three plus (three times three)'.

c.

[tʰa	mai	lə	fæn,	tæn	wo	mai	lə	ʂu]
他	买	了	饭,	但	我	买	了	书
H	L	0	F	F	R	L	0	H
ta	mai	le	fan	dan	wo	mai	le	shu
he	buy	ASP	rice	but	I	buy	ASP	book

'He bought rice, but I bought book(s).'

In choosing test sentences, factors that could affect duration were taken into consideration (Klatt 1975, 1976). The target syllables in the (a) sentences were the three numbers, indicated as an1, an2, and an3. The target syllables in the (b) sentences were also the three numbers, indicated as bn1, bn2, and bn3. (1a,b) and (2a,b) were modeled after the test sentences '(A plus E) times O' vs. 'A plus (E times O)' used by Streeter (1978); the parentheses indicate the structural ambiguities in each pair (which otherwise look identical). Phonetically, [tɛn] 'ten' in English and [sæn] 'three' in Chinese are the two digits that are closest in segmental and syllabic compositions, which was why they were chosen. In previous studies, it was found that the syntactic difference between pairs like (a) and (b) in (1) and (2) was conveyed primarily by means of duration (e.g. Lehiste 1973, Streeter 1978, Shen 1992). In the present case, an1 and bn2 are not pre-juncture,[i] whereas an2, an3, and bn3 are pre-juncture. Their durations are the focus of comparison.

The target syllable in the (c) sentences was [fæn], which was 'fan' in English and 'rice' in Chinese. The target syllables lay before a clause boundary and were subject to lengthening. Structurally (1c) and (2c) were similar with regard to syntax, the segments before and after the target syllables, the number and the kind of syllables before and after the target syllables, the number of stressed and unstressed syllables, etc.

2.1.2. Speakers and recording

12 American English speakers, six male and six female, were asked to read (1a-c). 12 Mandarin Chinese speakers, six male and six female, were asked to read (2a-c). The three sentences in each language were repeated five times and mixed in random orders; the resulting list contained 15 sentences. The list for English speakers was written in English. The list for Chinese speakers was written in Chinese characters. The ambiguities between (1a,b) and between (2a,b)

were indicated by parentheses in the reading list and were explained to the speakers in advance. The speakers were asked to express the ambiguities in a natural way (without reading the parentheses). All speakers found the ambiguities obvious and were confident that they could express them. Each speaker was given enough time to prepare before reading. The recording was made in ordinary rooms using a portable tape recorder and a microphone. All repetitions were used for measurements.

2.1.3. Measurements

Durational measurements were taken using the Mac Speech Lab. The placement of the cursors during the measurement was based on both the spectrogram and the waveform. Relevant items examined from each sentence are explained below.

2.2. Result

2.2.1. Sentences (1c) and (2c)

For (1c) and (2c) three items were measured: (i) the duration of the rime [æn] in the target syllables 'fan/rice', (ii) the duration of the pause after the target syllable, and (iii) the total duration of the sentence. For (ii), the beginning cursor was placed after [n] and the final cursor was placed at the release point of [b] or [t]. In other words, the closure duration (if any) of the following stop was included in the pause (if any).

The result of (i) is shown in Table 1. The average duration of the target rime [æn] was 354 ms in English and 192 ms in Chinese. A t-test shows that the difference between the two languages was highly significant.

The result of (ii) is shown in Table 2. When all speakers were included, the average pause duration in English was 112 ms and that in Chinese was 274 ms. A t-test showed that $p = 0.023$, which was significant at the 0.05 level. However, an examination of the data shows that the Chinese speaker C11 had excessively long pauses, at a mean of 860 ms, which was over two times the duration of the second longest mean (388 ms by C8). Apparently, C11 was the reason why the standard deviation for the Chinese speakers was nearly twice the value for the English speakers. A normality test (Lilliefors probability) showed that the distribution of the English speakers was normal ($p = 0.142$) but that of the Chinese speakers was borderline ($p = 0.053$). A graphic plot showed that C11 was an outlying value, which should be excluded. The resulting distribution of Chinese speakers (without C11) was normal (Lilliefors $p = 0.635$). The new mean for the Chinese speakers was 221 ms, with a standard deviation of 111 (which is comparable to 114 in English). A new t-test shows that $p = 0.008$, which indicates that the pause difference between English and Chinese was significant.

The difference in pause can also be seen in individual speakers. Of the 12 English speakers, 5 had a mean value below 50 ms. Recall that in the present measurement the duration of the pause contained the closure period of the following stop. If the closure of the following stop was properly excluded, those 5 English speakers would have a much smaller pause value, if any. In contrast, all the Chinese speakers had a mean pause value well above 50 ms. In other words, the pause in (1c) appears to be optional for English speakers but obligatory for Chinese speakers. The fact that the boundary pause is optional in English has been observed by Umeda (1975) and Wightman et al (1992). For example, according to Wightman et al (1992:1715), at 'intonational phrase' boundaries, a pause is found in just 23% of the cases.

In order to say that the difference between the target rimes in (1c) and (2c) was due to a difference in pre-juncture lengthening, one has to show that (i) Chinese and English do not differ in inherent syllable durations and (ii) tonal complexity was not a factor. For (i), previous studies have found that the average duration of an English syllable in read speech is about 200 ms

(Huggins 1964, Klatt 1976), and that of a Chinese syllable is also about 200 ms (Ao & Shih 1994). To check it again, the total durations of (1c) and (2c) were examined. For durational comparison, (1c) and (2c) were similar. First, each had nine syllables. Second, both sentences had 20 segments (counting diphthongs and tense vowels as single segments). Third, in the English sentence five of the syllables were unstressed, 'bought', 'a', 'but', 'bought', and 'a', so were there five unstressed syllables in the Chinese sentence, [mai lə] 'bought', [tæn] 'but', and [mai lə] 'bought'. Fourth, of the four stressed syllables in English, there was one high vowel and three non-high vowels; the same is true in the Chinese sentence. Finally, the two sentences were similar in meaning and syntax, so any semantic and phrasing factors should have had similar effects in both. Table 3 gives the total durations of (1c) and (2c), which show no significant difference. Thus, the difference between the target rimes was not due to inherent segment differences in the two languages.

Next consider tonal influence. For all Chinese speakers, the target rime carried its lexical tone HL (fall). In contrast, English speakers varied among four pitch contours. This is summarized in (3), where F = fall, R = rise, FR = fall-rise, and M = mid level.

(3) Pitch contours of the English target syllable 'fan' in (1c)

Contour	F (HL)	FR (HLH)	M	R (LH)	Total
Tokens	30	23	2	5	60

Let us focus on F and FR, which constituted the majority (88%). First, we compare the durations of F tokens and FR tokens, given in Table 4. The result shows that there was no significant difference between F tokens and FR tokens. Next we compare the F tokens in English with the F tokens in Chinese, which is shown in Table 5. Although all rimes in question had a falling tone, there was a significant durational difference between the two languages. Thus, the durational difference between English and Chinese was not the result of tonal complexity. Instead, the English rime was lengthened, consistently, whether there were three tones, as in FR (HLH), or just two, as in F (HL). This finding is consistent with those of previous studies.

2.2.2. Sentences (1a,b) and (2a,b)

The purpose of (1a,b) and (2a,b) was to compare pre-juncture rimes and non pre-juncture rimes in English and Chinese. The number bn1 was not treated consistently by the speakers: some had a pause after it (as well as a pause after 'plus'); others did not. So bn1 was not analyzed. Of the remaining five, an2, an3, and bn3 were pre-juncture, and an1 and bn2 were not. Their rimes were all analyzed.[ii]

The five English rimes are shown in Table 6. It can be seen from the average values that the pre-juncture rimes were longer than non pre-juncture rimes. In addition, among the three pre-juncture rimes, an2 was longer than the other two. Repeated Measures ANOVA was applied to determine whether the averaged differences were significant. The results are shown in Table 7. First, there was no significant difference between the two non pre-juncture rimes (an1 and bn2). Second, there was no significant difference between the two sentence final pre-juncture rimes (an3 and bn3). Third, between an2, which was pre-juncture but not sentence final, and an3 and bn3, which were pre-juncture and sentence final, there was a marginal difference ($p = 0.017$, which is significant for the 0.05 threshold but not significant for the 0.01 threshold). The fact that an2 was longer should not be surprising, since it lay at the main juncture for conveying the ambiguity between (1a) and (1b). Finally, a significant difference was found between non pre-juncture rimes (an1 and bn2) on the one hand and pre-juncture rimes (an2, an3, and bn3) on the other. This result reconfirms the well-known fact that stressed pre-juncture rimes were significantly longer than stressed non pre-juncture rimes in English.[iii]

Now consider Chinese. The five Chinese rimes are shown in Table 8 and analyzed in Table

9. There was no significant difference between non pre-juncture rimes, or among pre-juncture rimes. On the other hand, there was a significant difference between pre-juncture rimes and non pre-juncture rimes, as there was in English.[iv]

Next we compare English and Chinese. As seen in Table 6 and Table 8, the non pre-juncture rimes in English (an1 = 169 ms, bn2 = 168 ms) had similar durations as those in Chinese (an1 = 162 ms, bn2 = 170 ms). In addition, in both languages pre-juncture rimes were significantly lengthened. However, it can be seen that the size of lengthening was greater in English than in Chinese. To determine whether the between-language differences were significant, Repeated Measures ANOVA was performed on the averages of non pre-juncture rimes (an1 and bn2) and the averages of pre-juncture rimes (an2, an3, and bn3). The result is given in Table 10 and highlighted in Figure 1. First, in both English and Chinese a significant difference was found between pre-juncture and non pre-juncture rimes. Second, the difference between pre-juncture and non pre-juncture rimes was significantly dependent on the language. Third, non pre-juncture rimes did not differ in the two languages. Fourth, pre-juncture rimes in English were significantly longer than pre-juncture rimes in Chinese. Specifically, pre-juncture rimes lengthened by 75% in English, but just 28% in Chinese. These lengthening sizes are comparable with the findings of Umeda (1975), who found the increase in the English vowel to be 70%-100%, and the findings of Ao & Shih (1994), who found the increase in the Chinese syllable to be about 20%.

Let us now consider whether the extra lengthening in English was due to tonal complexity. The lexical tone for the Mandarin word [sæn] 'three' was H, which was maintained throughout. Of the 180 tokens of pre-juncture English rimes, 121 had F, 28 had FR, 2 had H, 17 had L, and 12 had R. The majority of them (F, H, L, and R, or 84%), therefore, did not need more than two moras. The Mandarin syllable [sæn] also has two moras (cf. Duanmu 1993). In addition, there is little durational difference between H, R, and F in Mandarin.[v] Thus, tonal complexity could not have contributed to the extra pre-juncture lengthening in English.

Finally, consider pauses. After an1 and bn2, the average silence durations were under 30 ms in both English and Chinese, which should belong to the closure of the following stop. Thus, there was no obvious pause after an1 and bn2. After an2, however, the average silence duration was above 100 ms in both English and Chinese, which means that there was some pause besides the closure of the following stop or affricate. This shows that both English and Chinese can use a pause in addition to pre-juncture lengthening at a syntactically and/or semantically contrastive juncture.

2.3. Summary

The present results are consistent with those of previous studies and give the following conclusions. First, both English and Chinese are subject to pre-juncture lengthening. Second, English exhibits significantly more pre-juncture lengthening than Chinese does. Third, Chinese has a greater tendency to use pauses at syntactic junctures than English does. It will be helpful, of course, if more experiments are carried out on other rime pairs in English and Chinese. But the consistency in the results obtained so far provides good indication that further results will point to the same way. In what follows I will assume that the present conclusions are essentially correct.

3. Phonological Analysis

Let us now explore the reason for the difference in pre-juncture lengthening between English and Chinese. First, consider the idea that Chinese lacks pre-juncture lengthening because it is a tone language in which every regular syllable has a lexical tone. Since a tonal contour is a function of both pitch height and time, changing the duration of a syllable will presumably change its tonal contour. And since tonal contour is a distinctive property of every regular Chinese syllable, it is plausible that changes in tone are generally avoided. However, this proposal has three shortcomings. First, it is not the case that there is no pre-juncture lengthening in Chinese; we have seen that there is significant pre-juncture lengthening in Chinese, just as there is English. Second, although lengthening a syllable may change its tonal contour, the distinctions among tonal categories need not be lost. For example, Mandarin has four tones on full syllables, which are often described as high level, high rise, fall-rise, and fall. If we interpret them as H, MH, MLH, and HL, and assume that the lengthened syllable has three moras, the resulting tones, over the moras, will be HHH, MHH, MLH, and HLL, where the four categories remain distinct. Third, changes in tone do occur in Chinese, despite the loss of tonal contrast. For example, the well-known Mandarin Third Tone Sandhi changes the Third tone into the Second, which can introduce ambiguities. For example, [mai3] 'buy' has the Third tone and [mai2] 'bury' has the Second tone. In [mai2 ma3] 'to buy a horse', [mai3] has changed to the Second tone because of the following Third tone in [ma3] 'horse'. As a result, the expression sounds identical to 'to bury a horse'. Thus, the fact that a pre-juncture Chinese syllable lengthens less cannot be attributed to the need to preserve tonal distinctions.

What else then is the explanation? I suggest that pre-juncture lengthening is not a homogeneous phenomenon, but is sensitive to both phonetics and phonology. Let us call the former phonetic lengthening and the latter phonological lengthening.[vi] The former is probably due to a 'slow-down' tendency at the end of a structure.[vii] I assume then that both English and Chinese are subject to phonetic lengthening, in accordance with the fact that pre-juncture rimes were longer than non pre-juncture rimes in both languages. However, unlike phonetic lengthening, phonological lengthening is not universal, but requires specific conditions. I will argue that these conditions are met in English but not in Chinese. Consequently, phonological lengthening occurs in English but not in Chinese, which leads to the fact that English pre-juncture rimes are longer than those in Chinese.

Consider first the motivation for phonological lengthening in English. I propose that it is the need to satisfy foot binarity. A fundamental property of stress and rhythm is the alternation between strong and weak beats. To realize such an alternation there needs to be a minimum of two beats (syllables). In phonology this is known as foot binarity. In earlier literature foot binarity was thought to be a soft requirement, but recent research suggests that it is stronger than has been realized. For example, according to Halle & Vergnaud (1987), English word stress usually falls on a binary foot, yet in monosyllabic words it falls on a monosyllabic foot. In contrast, Burzio (1994) argues that all English feet are at least binary. What used to be thought as monosyllabic feet, such as 'fan' and 'bee', are in fact disyllabic feet (fan.nø) and (bee.ø), where the second syllable has a phonetically null vowel [ø]. Similarly, because English has left-headed feet, a stressed final syllable in such words as 'Tennessee' forms a monosyllabic foot in Halle & Vergnaud's analysis, but a disyllabic binary foot 'Tennes(see.ø)' in Burzio's analysis. Although null elements may appear counter-intuitive, they have been proposed in the literature. For example, silent beats in phonology have been proposed by Abercrombie (1971), Liberman (1975), and Selkirk (1984). Nevertheless, there remain two questions about the null vowel. First, what is the motivation for it? Second, how is it realized? For the first question, Burzio suggests a cross-linguistic constraint that all words must end in a vowel. This may explain why 'fan' is analyzed as (fan.nø) (ignoring consonant gemination), but not why 'bee' should be analyzed as (bee.ø).[viii] For the second question, one would expect the null vowel to be realized as a silent pause, perhaps. But there is no such evidence. For example, according to Umeda (1975), stressed pre-juncture English rimes, which for Burzio should end in a null vowel, are more often followed by no pause than by a pause in read speech. A better analysis, therefore, is not to posit the null vowel for English, yet still preserve Burzio's key insights. I suggest that instead of analyzing 'fan' as [fæn.nø] and 'bee' as [bi.ø], they be analyzed as [fææ.n] or [fæ.æn] and [bi.i]. In other words, they are disyllabic, but without the null vowel. Because the disyllabic forms contain more segment slots, they are expected to be longer than non pre-juncture [fæn] and [bi]. This analysis preserves the binary foot minimum, a key in Burzio's analysis, and avoids the difficulties with the null vowel. In addition, this analysis accounts for the fact that a stressed pre-juncture English syllable is significantly lengthened (beyond the amount of phonetic lengthening observed in Chinese), as shown in the present experiment and previous literature.

Let us now turn to the lack of phonological lengthening in Chinese. First, consider stress in Mandarin. Unlike English, where many words are polysyllabic and where intuition for stress is reasonably clear, Chinese is a monosyllabic language in which the question does not arise as to where stress falls on a simple word. There is, instead, a fairly clear distinction between full syllables, which are usually content words, and unstressed (toneless) syllables, which are usually functional words or particles. The question of stress does arise in compounds (and phrases) made of two or more full syllables, but as Chao (1968:38) points out, whereas there is a good degree of agreement on which syllables are unstressed, opinions differ in regard to relative degrees of stress among full syllables of a compound. Still, some linguists made an attempt to capture the nuances. For example, Chao (1968:29) suggests that in Mandarin 'a two-syllable compound or phrase will have a slightly greater stress on the second syllable.' A similarly proposal is made by Lin et al (1984). Now if Mandarin has right-headed feet, then perhaps one ought not to not expect the same phonological lengthening as in English, which has left-headed feet.

But there is some evidence that Mandarin has left-headed stress, just as English does. Most people agree that duration is a primary indication of stress in Chinese (Chao 1968, Lin et al 1984, Yan & Lin 1988, Wang & Wang 1993). According to Lin et al (1984), the second syllable of a disyllabic Mandarin compound is longer than the first, and according to Yan & Lin (1988), the final syllable of a trisyllabic Mandarin compound is longer than the other two. However, according to Wang & Wang (1993), the first syllable is longer in both disyllabic and trisyllabic compound. One may wonder why there is contradicting evidence. A close look at the methodology suggests that Wang & Wang's result is more reliable. In the experiments of Lin et al (1984:66) and Yan & Lin (1988:228), target words were read in isolation. In Wang & Wang's experiments, target words were read in a carrier sentence. It is likely, then, that the longer duration of the final syllable reported by Lin et al (1984) and Yan & Lin (1988) was due to final lengthening, and not to final stress. Further evidence for left-headed stress in Mandarin compounds comes from the distribution of unstressed (toneless) syllables. As Lin (1994) points out, in disyllabic compounds there are many cases where the second syllable is unstressed, but there is no case where the first syllable is unstressed. Similarly, in four-syllable compounds there are many cases where the second and third syllables are unstressed, but there is no case where the first and the third syllables are unstressed.

It appears then that Mandarin has left-headed feet after all, just as English does. But then we are back to the original question again: why is there more pre-juncture lengthening in English than in Mandarin? I suggest that the answer lies in syllable inventory. English has a wide range of syllable structures. Preceding the vowel there can be from zero to three consonants. The vowel itself can be short (lax), long (tense), or a diphthong. After the vowel there can be up to four consonants. Although there are some constraints on the English syllable, such as sonority (for onset and coda clusters) and place of articulation (for coda clusters), there is still a large inventory of syllable types. In contrast, syllable types in Chinese is radically limited. In fact, according to Duanmu (1993), there is just one structure, CVX, for all regular syllables in Mandarin (a small number of lexically weak syllables have the structure CV).[ix] Similar proposals are made by Yip (1992) and Wang (1993). The C represents a consonant or a glide (the consonant can also have a secondary articulation, which is traditionally written as a separate glide). The V represents a short vowel, the first mora of a diphthong, or the first mora of a long vowel. The X represents either the second mora of a long vowel, or the second mora of a diphthong, or a nasal coda. The richness of the English syllable inventory allows it to turn a monosyllable into a disyllable easily (when called upon by foot binarity). For example, [tɛn] 'ten' can turn into [tɛ.ɛn], [tɛɛ.n], or [tɛɛ.ɛn], without creating ill-formed syllable types. In Chinese, however, such freedom is not available, because CVX is the only syllable structure. For example, one cannot turn [sæn] 'three' into [sæ.æn], [sææ.n], or [sææ.æn]; in the first case both syllables are ill-formed and in the second and third the second syllable is ill-formed. Thus, although a stressed pre-juncture syllable is subject to foot binarity in both English and Chinese, English can satisfy it by lengthening, but Chinese cannot.

Lengthening is not the only way to satisfy foot binarity, however. Ironically, while the idea of null syllables did not seem appropriate for English, it works fine for Chinese. In particular, I suggest that a stressed pre-juncture syllable in Chinese is followed by a silent beat, and together they form a binary foot. This analysis directly accounts for the two differences between English and Chinese, which have so far remained unexplained: first, English had greater pre-juncture lengthening than Chinese, and second, the pause after a stressed pre-juncture syllable is often absent in English but rarely absent in Chinese.

To summarize, I have proposed that both phonetics and phonology can influence pre-juncture lengthening. Both English and Mandarin undergo phonetic pre-juncture lengthening. However, they differ in phonological pre-juncture lengthening as follows. Both English and Mandarin have left-headed feet and both observe foot binarity. A stressed pre-juncture syllable does not form a binary foot and repair strategies are needed. In English the pre-juncture syllable is turned into two syllables through lengthening, made possible by the flexibility in syllable structure. In Mandarin the same lengthening is not available, owing to the rigid CVX syllable structure. Instead, Mandarin satisfies foot binarity by using a silent beat (pause) after a stressed pre-juncture syllable. This analysis accounts for the three phonetic results reported in the section 2: (i) there is pre-juncture lengthening in both English and Chinese, (ii) English shows more pre-juncture lengthening than Mandarin, and (iii) a pause after the pre-juncture syllable is optional in English but required in Mandarin.

4. Conclusions

I have given experimental results that English shows more lengthening in a stressed pre-juncture syllable than Chinese. To explain the difference, I propose that pre-juncture lengthening is influenced by both phonetics and phonology. Phonetic lengthening is present in both English and Chinese. Phonological lengthening is present in English but not in Chinese. Phonological lengthening in English is motivated by the need to satisfy foot binarity; in particular, a stressed final syllable does not form a binary foot, and to satisfy foot binarity, it will split into two syllables. This analysis is similar (but not identical) to the original insight of Burzio (1994). Thus, [fæn] 'fan' is turned into [fæ.æn] or [fææ.n]. What appears to be syllable lengthening, therefore, is in fact the result of syllable split. Like English, Chinese also obeys foot binarity, but owing to its rigid syllable structure, syllable split is not available. In particular, CVX is the only structure for full Mandarin syllables (Duanmu 1993); a syllable like [fæn] 'rice' cannot be split into [fæ.æn], [fææ.n], or [fææ.æn]; in the first both syllables are ill-formed, and in the second and third the second syllable is ill-formed. As a result, Chinese needs to satisfy foot binarity in other ways, such as using a pause (or a filler word) as the second syllable of a foot.

The present analysis raises a range of questions. For example, is there independent evidence that a lengthened English syllable counts as two syllables? Is there independent evidence that a pause can count as a syllable in Chinese? Can the present analysis be extended to other languages? Will a stressed monosyllabic word be lengthened in nonfinal positions? Are there other consequences of the difference in pre-juncture lengthening between English and Chinese? Are there exceptions to the present claim that Chinese syllables can not be lengthened? Is phonological pre-juncture lengthening categorical, i.e. either present or absent? For lack of space, these issues will not be addressed here.

Notes

*For various comments and help, thanks to Pam Beddor, Lisa Cheng, Jennifer Cole, Andre Cooper, Shengli Feng, Brenda Gillespie, Yong He, Jim Huang, Yen-hwei Lin, James Myers, Jane Tsay, Kathy Welch, Moira Yip, Shangyang Zhao, and audiences at University of California Irvine, Ohio State University, NACCL 8, and Center for Chinese Studies at University of Michigan.

[i]Strictly speaking, an1 and bn2 also lay before a word juncture and an NP juncture. However, the main junctures in question, namely, that after an2 and that before bn2, made word and phrase junctures less important.

[ii]Three sentence tokens were mistakenly read: the Chinese speaker C7 read (4a) as (4b) on the fifth repetition, and the Chinese speaker C11 read (4a) as (4b) on the second and third repetitions. These three tokens were discarded. The place of each discarded token was filled by the average values of the good repetitions of the same sentence by the same speaker.

[iii]The statisticians I consulted had differing opinions as to whether the comparisons in Table 7 should be adjusted by the Bonferroni factor. The Bonferroni adjustment is sometimes applied in multiple comparisons, by which the significance level is divided by the number of comparisons made. In Table 7, four comparisons were made. If one assumes 0.05 to be the significance level, then the adjusted significance level will be 0.0125. Thus, for example, the difference between an2 vs. an3-bn3 ($p = 0.017$) is significant for the unadjusted significance level of 0.05, but not significant for the adjusted significance level of 0.0125. An alternative way of analyzing the data in Table 6 is to compare every rime with everyone else by paired t-test (10 pairs in all), and then adjust the results with the Bonferroni factor of 10 (i.e. using either 0.005 or 0.001 as the significance level). The results are similar to those in Table 7.

[iv]It can be seen that the Bonferroni adjustment will not alter the result in Table 9.

[v]Reported variations among tone 1, tone 2, and tone 4 are small and inconsistent. For example, according to Luo & Wang (1957: 127), tone 2 is 7% longer than tone 4, but according to Howie (1976: 20), tone 2 is just 2% longer than tone 4. Similarly, according to Luo & Wang tone 1 is 2.6% longer than tone 4, but according to Howie tone 4 is 6% longer than tone 1.

[vi]Other phonological lengthening phenomena such as compensatory lengthening are not discussed here.

[vii]According to Wightman et al (1992), pre-juncture lengthening in English is restricted to the final rime only.

[viii]Burzio (p.c.) suggests that the null vowel in words like '(bee.ø)' and 'kanga(roo.ø)' is forced purely by the prohibition against a unary foot.

[ix]The phonetic transcription in (2) reflects the traditional view of Chinese syllables rather than that of Duanmu's. Duanmu (1993) also suggests that another group of Chinese dialects, notably Shanghai and some of its neighbors in the Wu dialect family, have a fixed syllable structure CV. If pre-juncture lengthening is related to syllable structure, this group should behave like Mandarin and unlike English, in that they resist pre-juncture lengthening and tend to have pauses at junctures.

References

Abercrombie, David. 1971. Some functions of silent stress. In Edinburgh studies in English and Scots, ed. A.J. Aitken, Angus McIntosh, and Hermann Pálsson, 147-156. London: Longman.

Ao, Benjamin, and Chilin Shih. 1994. Contextual variations of segmental duration in Mandarin Chinese. Paper presented at NACCL 6, University of Southern California.

Berkovits, Rochele. 1991. The effect of speaking rate on evidence for utterance-final lengthening. Phonetica 48.1: 57-66.

Burzio, Luigi. 1994. Principles of English stress. Cambridge, UK: CUP.

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

Delattre, Pierre. 1965. Comparing the phonetic features of English, French, German, and Spanish: An interim report. London: Harrap.

Duanmu, San. 1993. Rime length, stress, and association domains. Journal of East Asian Linguistics 2.1:1-44.

Halle, Morris, and Jean-Roger Vergnaud. 1987. An essay on stress. Cambridge, Mass.: MIT Press.

Howie, John. 1976. An acoustic study of Mandarin tones and vowels. London, England: Cambridge University Press.

Huggins, A.W.F. 1964. Distortion of the temporal pattern of speech: interruption and alternation. *Journal of the Acoustical Society of America* 36.10:1055-1064.

Klatt, Dennis H. 1975. Vowel lengthening is syntactically determined in a connected discourse. Journal of Phonetics 3.3:129-140.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59.5:1208-1221.

Lehiste, Ilse. 1973. Phonetic disambiguation of syntactic ambiguity. Glossa 7.2:107-121.

Liberman, Mark. 1975. The intonational system of English. Doctoral dissertation, MIT.

Lin, Hua. 1994. Against domain-final stress in Mandarin. Paper presented at NACCL 6, University of Southern California.

Lin, Maochan , Jing-zhu Yan, and Guo-hua Sun. 1984. Beijinghua lianzizu zhengchang zhongyin de chubu shiyan [Preliminary experiments on the normal stress in Beijing disyllables]. Fangyan 1984.1:57-73.

Luo, Chang-Pei, and Jun Wang. 1957. Putong yuyinxue gangyao [Outline of general phonetics].

Peking: Kexue Chubanshe. Reprinted in 1981 by Shangwu Yinshuguan.

Selkirk, Elisabeth. 1984. Phonology and syntax: The relation between sound and structure, Cambridge, Massachusetts: MIT Press.

Shen, Susan. 1992. A pilot study on the relation between the temporal and syntactic structures in Mandarin. Journal of the International Phonetic Association 22.1-2:35-43.

Streeter, Lynn Anne. 1978. Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America* 64:1582-1592.

Umeda, Noriko. 1975. Vowel duration in American English. *Journal of the Acoustical Society of America* 58.2:434-445.

Wang, Jing, and Lijia Wang. 1993. Putonghua duo yinjie ci yinjie shi chang fenbu moshi. Zhongguo Yuwen 1993.2 (233): 112-116.

Wightman, Colin W., Mari Ostendorf, Stefanie Shattuck-Hufnagel, and Patti Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. Journal of the Acoustical Society of America 91.3:1707-17.

Yan, Jingzhu, and Maocan Lin. 1988. Beijinghua sanzizu zhongyin de shengxue biaoxian [Acoustic characteristics of the stress in Beijing trisyllables]. Fangyan 1988.3: 227-237.

Yip, Moira. 1992. Prosodic morphology in four Chinese dialects. Journal of East Asian Linguistics 1.1:1-35.

	1	2	3	4	5	Mean	
E1	355	350	293	374	346	344	
E2	470	517	531	470	484	494	
E3	394	403	422	403	403	405	
E4	410	379	418	389	363	392	
E5	403	379	389	442	379	398	
E6	398	384	312	278	292	333	
E7	307	307	326	252	307	300	
E8	323	265	239	235	235	259	
E9	302	322	346	298	269	307	
E10	331	274	331	379	384	340	
E11	341	396	413	331	360	368	
E12	322	336	298	264	289	302	SD = 67.6
C1	211	182	173	254	178	200	
C2	202	154	187	192	148	177	
C3	154	144	202	216	114	166	
C4	197	222	235	208	221	217	
C5	216	254	208	208	195	216	
C6	240	226	221	206	192	217	
C7	178	222	173	182	221	195	
C8	168	178	163	187	163	172	
C9	144	139	178	110	111	136	
C10	178	208	187	187	188	190	
C11	222	208	188	228	206	210	
C12	205	230	198	205	215	211	SD = 32.2

Average English = 354 (ms)
Chinese = 192 (ms)
T-test: p = 0.000

Table 1: Durations in ms of the target rime [æn] in the English (1c) and the Chinese (2c). E1-E12 represent 12 English speakers. C1-C12 represent 12 Chinese speakers. 1-5 on the top row represent five repetitions of each speaker. The standard deviation (SD) in each language was based on the repetitions by all speakers. A t-test, based on the means of the speakers, shows p = 0.000, which means that the between-language difference in duration was significant.

	1	2	3	4	5	Mean	
E1	245	120	58	58	192	135	
E2	67	222	74	484	276	225	
E3	250	149	182	269	307	231	
E4	155	173	274	154	336	218	
E5	206	0	58	274	163	140	
E6	29	24	0	0	0	11	
E7	53	34	29	24	24	33	
E8	50	44	30	34	40	40	
E9	0	10	0	0	0	2	
E10	10	0	0	197	0	41	
E11	43	262	72	77	82	107	
E12	0	58	226	206	316	161	SD = 114
C1	274	206	187	96	144	181	
C2	91	302	240	221	208	212	
C3	182	120	86	72	138	120	
C4	235	81	158	225	139	168	
C5	211	178	121	121	108	148	
C6	250	216	254	173	216	222	
C7	302	74	346	331	125	236	
C8	514	451	350	274	350	388	
C9	307	264	374	307	302	311	
C10	490	403	336	283	235	349	
(C11	974	806	948	813	758	860)	
C12	185	96	77	57	70	97	SD = 111

Average English = 112 (ms)
Chinese = 221 (ms)

T-test $p = 0.008$

Table 2: Durations in ms of the pause after the target syllable in the English (1c) and the Chinese (2c). E1-E12 represent 12 English speakers. C1-C12 represent 12 Chinese speakers. 1-5 on the top row represent five repetitions of each speaker. A check on distribution normality shows that C11 was an outlying value, which was excluded from the calculations. See text for more discussion.

	1	2	3	4	5	Mean	
E1	1992	1882	1790	1887	1867	1884	
E2	2507	2614	2722	2789	2607	2648	
E3	2102	2078	2002	2030	2030	2048	
E4	2251	2208	2261	2136	2204	2212	
E5	1978	1757	1886	2198	1978	1959	
E6	1886	1882	1651	1584	1636	1728	
E7	1800	1848	1723	1622	1704	1739	
E8	1499	1441	1327	1294	1294	1371	
E9	1829	1992	1958	1920	2098	1959	
E10	1810	1670	1690	1978	1642	1758	
E11	1925	2137	1853	1781	1824	1904	
E12	1872	1776	1901	1685	1636	1774	SD = 315
C1	1886	1858	1781	1694	1675	1779	
C2	1814	1978	1838	1819	1646	1819	
C3	1930	1853	1690	1656	1606	1747	
C4	1853	1667	1786	1835	1694	1767	
C5	1843	1843	1667	1616	1593	1712	
C6	2357	2107	2150	1997	1862	2095	
C7	1896	1643	1973	1963	1656	1826	
C8	2170	2112	2002	1963	1939	2037	
C9	1627	1589	1771	1598	1562	1629	
C10	2174	2076	1954	1891	1744	1968	
C11	2748	2527	2661	2560	2376	2574	
C12	1680	1555	1509	1542	1485	1554	SD = 285

Average English = 1915 (ms)
Chinese = 1876 (ms)

T-test: $p = 0.741$

Table 3: Total durations in ms of the English sentence (1c) and the Chinese sentence (2c). E1-E12 represent 12 English speakers. C1-C12 represent 12 Chinese speakers. 1-5 on the top row represent five repetitions of each speaker. The standard deviation (SD) in each language was based on the repetitions by all speakers. A t-test, based on the means of the speakers, shows that that the difference in total durations between (1c) and (2c) was not significant.

FR (fall-rise)

	1	2	3	4	5	Mean	
E1					346	346	
E2	470	517	531	470	484	494	
E3	394	403	422	403	403	405	
E4							
E5							
E6		384	312			348	
E7		307	326	252	307	298	
E8		265		235		250	
E9							
E10				379		379	
E11			413	331	360	368	
E12							SD = 78.5

F (fall)

E1	355	350	293	374		343	
E2							
E3							
E4	410	379	418	389	363	392	
E5	403	379	389	442	379	398	
E6	398			278	292	323	
E7							
E8	323		239		235	266	
E9							
E10		274	331		384	330	
E11	341	396				369	
E12	322	336	298	264	289	302	SD = 53.0

Average FR = 361 (ms)
F = 340 (ms)

t-test $p = 0.505$

Table 4: Durations in ms of the target rime [æn] in the English (1c), grouped by pitch contours. The upper set shows durations for the contour FR (fall-rise). The lower set shows durations for the contour F (fall). E1-E12 represent 12 English speakers. 1-5 on the top row represent five repetitions of each speaker. A t-test on the means of the speakers shows no significant difference between the two pitch contour groups.

	1	2	3	4	5	Mean	
E1	355	350	293	374		343	
E2							
E3							
E4	410	379	418	389	363	392	
E5	403	379	389	442	379	398	
E6	398			278	292	323	
E7							
E8	323		239		235	266	
E9							
E10		274	331		384	330	
E11	341	396				369	
E12	322	336	298	264	289	302	SD = 53.0
C1	211	182	173	254	178	200	
C2	202	154	187	192	148	177	
C3	154	144	202	216	114	166	
C4	197	222	235	208	221	217	
C5	216	254	208	208	195	216	
C6	240	226	221	206	192	217	
C7	178	222	173	182	221	195	
C8	168	178	163	187	163	172	
C9	144	139	178	110	111	136	
C10	178	208	187	187	188	190	
C11	222	208	188	228	206	210	
C12	205	230	198	205	215	211	SD = 32.2

Average English = 340 (ms)
Chinese = 192 (ms)

t-test $p = 0.000$

Table 5: Durations in ms of the target rime [æn] in the English (1c) and the Chinese (2c). For English speakers (E1-E12), only those tokens whose pitch contours were F (fall) were included. The pitch contours of all tokens by Chinese speakers (C1-C12) were F. 1-5 on the top row represent five repetitions of each speaker. The t-test was based on the means of the speakers. Although all rimes had a falling pitch contour, there was a significant durational difference between the two languages.

	Non pre-juncture		Pre-juncture			
EN	an1	bn2	an2	an3	bn3	
1	178	155	314	306	285	
2	293	285	433	376	387	
3	167	150	354	294	307	
4	170	177	331	253	270	
5	124	137	343	342	335	
6	137	158	246	315	345	
7	196	204	307	206	213	
8	134	119	235	190	179	
9	185	182	383	271	255	
10	156	162	389	319	287	
11	150	155	290	279	298	
12	136	135	265	210	206	
SD	47	44	65	64	67	
All	169	168	324	280	281	(ms)

Table 6: Durations in ms of the English rime [ɛn] in various positions in (1a) and (1b), read by 12 English speakers. Each entry for a speaker was the average of 5 repetitions. The standard deviation (SD) was based on the repetitions. The data are analyzed by Repeated Measures ANOVA, which is shown in Table 7.

Comparisons of English Rimes under Repeated Measures ANOVA

Between non pre-juncture rimes (an1 vs bn2)

$p = 0.880$

Between sentence final pre-juncture rimes (an3 vs bn3)

$p = 0.927$

Nonfinal pre-juncture (an2) vs. final pre-juncture (an3 and bn3)

$p = 0.017$

Non pre-juncture (an1 and bn2) vs. pre-juncture (an2, an3, and bn3)

$p = 0.000$

Table 7: Comparisons under Repeated Measures ANOA of the five English rimes (data in Table 6). First, there was no significant difference between the non pre-juncture rimes (an1 and bn2). Second, there was no significant difference between the sentence final pre-juncture rimes (an3 and bn3). Third, there was a marginal difference between an2, which was pre-juncture but not sentence final, and an3 and bn3, which were pre-juncture and sentence final. Finally, there was a significant difference between non-prejuncture rimes (an1 and bn2) on the one hand and pre-juncture rimes (an2, an3, and bn3) on the other.

	Non pre-juncture		Pre-juncture		
CH	an1	bn2	an2	an3	bn3
1	152	163	196	191	189
2	158	153	189	213	192
3	166	180	216	179	186
4	159	152	247	230	210
5	153	162	228	233	218
6	201	220	235	257	278
7	176	174	227	192	177
8	140	171	186	164	188
9	128	128	160	180	180
10	158	173	255	222	209
11	169	143	230	232	234
12	179	216	251	235	214
SD	21	28	38	33	31
All	162	170	218	211	206

Table 8: Durations in ms of the Chinese rime [æn] in various positions in (2a) and (2b), read by 12 Chinese speakers. Each entry for a speaker was the average of 5 repetitions. The standard deviation (SD) was based on the repetitions. The data are analyzed by Repeated Measures ANOVA, which is shown in Table 7.

Comparisons of Chinese Rimes under Repeated Measures ANOVA

Between non pre-juncture rimes (an1 vs. bn2)

$p = 0.138$

Between sentence final pre-juncture rimes (an3 vs. bn3)

$p = 0.353$

Nonfinal pre-juncture (an2) vs. final prejuncture (an3 and bn3)

$p = 0.188$

Non-prejuncture (an1 and bn2) vs. pre-juncture (an2, an3, and bn3)

$p = 0.000$

Table 9: Comparisons under Repeated Measures ANOA of the five Chinese rimes (data in Table 8). First, there was no significant difference between the non pre-juncture rimes (an1 and bn2). Second, there was no significant difference between the sentence final pre-juncture rimes (an3 and bn3). Third, there was no difference between an2, which was pre-juncture but not sentence final, and an3 and bn3, which were pre-juncture and sentence final. Finally, there was a significant difference between non-prejuncture rimes (an1 and bn2) on the one hand and pre-juncture rimes (an2, an3, and bn3) on the other.

	English		Chinese	
Spkr	NPJ	PJ	NPJ	PJ
1	167	302	158	192
2	289	399	156	198
3	159	318	173	194
4	174	285	156	229
5	131	340	158	226
6	148	302	211	257
7	200	242	175	199
8	127	201	156	179
9	184	303	128	173
10	159	332	166	229
11	153	289	156	232
12	136	227	198	233
All	**169**	**295**	**166**	**212**

Repeated Measures ANOVA

Cross-language NPJ vs. PJ:	$p = 0.000$
Interaction between language & lengthening:	$p = 0.000$
Cross-language NPJ:	$p = 0.833$
Cross-language PJ:	$p = 0.000$

Table 10: Pre-juncture (PJ) and non pre-juncture (NPJ) rimes in English and Chinese, with 12 speakers in each language. First, there is a difference between PJ and NPJ rimes across the two languages. Second, the difference between NPJ and PJ rimes is dependent on the language. Third, NPJ rimes do not differ in English and Chinese. Fourth, PJ rimes do differ in English and Chinese.

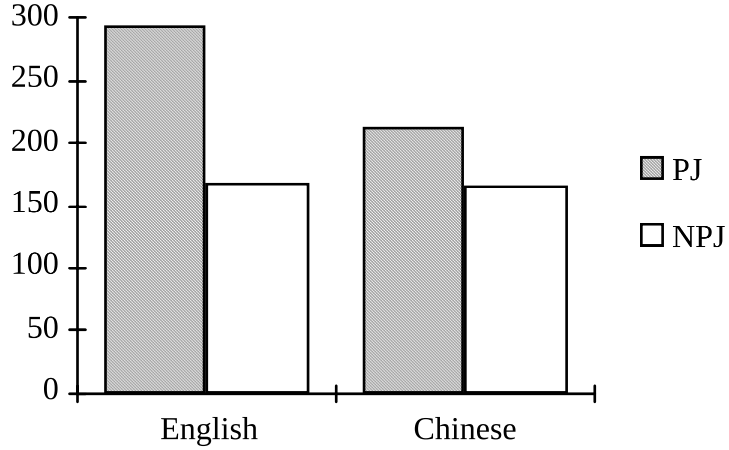

Figure 1: Average durations in ms of pre-juncture rimes (PJ) and non pre-juncture rimes (NPJ) in English and Chinese (see Table 10). NPJ rimes are similar in the two languages, but PJ rimes are lengthened significantly more in English than in Chinese.