F_0 PEAK DELAY: WHEN, WHERE, AND WHY IT OCCURS

Yi Xu
Northwestern University, Evanston, IL, USA

ABSTRACT

Peak delay refers to the phenomenon that an F_0 peak sometimes occurs after the syllable it is associated with. This study investigates the relationship between tone, speaking rate and peak delay in Mandarin. Sentences containing H, R, or weakened H (h) were recorded at normal, fast and slow speaking rates. At normal rate, peak delay occurred regularly in both R and h but only occasionally in H; at slow rate, peak delay continued to occur regularly in R but only occasionally in h and rarely in H; at fast rate, peak delay occurred not only regularly in R and h, but also frequently in H. Peak-alignment analyses revealed that peak delay tended to occur whenever F_0 rose sharply near the syllable offset. The results were interpreted as indicating that peak delay is due to an articulatory constraint that limits how fast the larynx can reverse the direction of pitch movement.

1. INTRODUCTION

If a tone, pitch accent, stress or focal prominence is carried by a syllable, by the simplest assumption, the F_0 peak associated with it should occur somewhere inside that syllable. However, it has been noted in recent studies [1, 6, 7, 8, 10, 11, 14, 15] that the peak often occurs after the syllable that carries the tone, pitch accent, stress or focal prominence. The present paper refers to this phenomenon as 'peak delay'.[1]

Peak delay has been reported, explicitly or implicitly, for a number of languages: English [11], Korean [7], Mexican Spanish [10], Oneida [6], Greek [1], Chichewa [8], and Mandarin [14, 15]. There has been little discussion, however, on why and how peak delay occurs. To investigate possible underlying mechanisms for peak delay, it is necessary to examine first why F_0 peaks occur and how they are aligned with segmental units in general.

1.1. F_0 peaks

Peaks can be easily observed in the F_0 tracings of speech utterances. It is therefore natural, to a first approximation, to assume that each peak is somehow directly related or even equivalent to some kind of underlying pitch target. Based on this assumption, F_0 contours can be viewed as the natural consequences of interpolation between peaks and valleys [9]. Indeed it has been shown that various Swedish dialects can be effectively classified in terms of the number and location of "turning points" in the F_0 curves of their pitch accents [3]. It has also been argued that turning point can be used to distinguish Mandarin tones [4, 5]. However, as found in recent studies [13, 15], in Mandarin, none of the lexical tones has to always have a peak. Rather, a tone may have a peak only when it is in an appropriate tonal or prosodic context. For example, H may not have a peak when surrounded by H; and R may not have a peak when followed by H. What is more consistent about a tone across various tonal contexts, speaking rates, and focus conditions in Mandarin is the F_0 contour in the later portion of its host syllable.

1.2. Peak location

Even with the right tonal context, the relative location of the peak may vary depending on the inherent properties of the tonal target. In Mandarin, it has been found that F_0 peaks in different tones have distinct alignment patterns [15]. In F the peak occurs near the center of the syllable (when preceded by L or F); in H the peak occurs close to but before the offset of the host syllable (when preceded by L or F and followed by L); and in R the peak usually occurs *after*, but still close to the syllable offset (when followed by L).

To account for these alignment patterns, Xu and Wang [16] propose that F_0 contour alignment patterns in Mandarin can be more easily explained if tonal targets are assumed to be aligned to the entire syllable rather than to any particular portion of the syllable. Under this assumption, F and R have diametrically different peak alignments because the falling contour in F necessitates a relatively early high point in a syllable while the rising contour in R necessitates a relatively late high point. Likewise, peak alignment in H falls between R and F because its high point does not have to be either extremely early or extremely late. In other words, the three distinct peak alignment patterns of F, R and H may be directly attributable to the different underlying phonological properties of these tones. Also under this assumption, the implementation of the tonal target should start at the syllable onset and end at the syllable offset. Since it takes time for the larynx to effectively implement a tonal target, F_0 contour alignments in all tones would appear somewhat late in a syllable, as has been found in recent studies [13, 15]. As a result, the underlying target of a tone should always be approximated the fullest in the final portion of a syllable [16].

1.3. Peak delay

Based on the above account, since R has a rising target, the sharpest F_0 rise in R should be achieved just before the syllable offset. When R is followed by L, F_0 should start to fall as soon as the L-carrying syllable starts (because the implementation of L starts there). However, if it takes time for the larynx to terminate the rise, the turning point in F_0, i.e., the peak, would occur somewhere after the boundary between the two syllables that carry the RL sequence. In contrast, in H the implementation of a static high target after L should cause F_0 to rise and then level off within the syllable. It is then less likely that the peak would occur after the syllable boundary even when H is followed by L. This has indeed been found to be the case in Mandarin [15].

A natural prediction of such an account is that peak delay should occur whenever there is a sharp F_0 rise just before the syllable offset, regardless of the underlying cause of the F_0 rise. For example, in a LHL sequence, if the duration of the H-carrying syllable is substantially shortened for some reason, the F_0 rise in H may be pushed close to the end of the syllable, just as in R. In such a case, would peak delay also occur, just as in R? An experiment was therefore carried out to explore this possibility.

2. FORCING PEAK DELAY TO OCCUR IN H — AN EXPERIMENT

2.1. Experimental Design

2.1.1. Stimuli. The basic design of the experiment is to manipulate syllable duration, target tone, and tonal context to see how F_0 peak alignment is affected. Syllable duration was controlled mainly by changing speaking rate. Three speaking rates were used, normal, fast and slow. To control tone, three tones conditions were used: R, H and weakened H. A weakened H occurs in Mandarin in the middle syllable of a verb phrase such as ‘mǎyīmǎ’ ‘to stack up’. The middle syllable in this kind of phrase structure is prosodically weak and usually has shortened duration.[2] The tonal context, L__L, was held constant across the three tone conditions to ensure the occurrence of an F_0 peak in or near the middle syllable. Twelve sentences were used in the experiment. They are divided into three groups. Group 1 contains the LHL sequence; Group 2 contains the LhL sequence, where h denotes a weakened H; and Group 3 contains the LRL sequence. One example sentence from each group is shown in Table 1.

(1) LHL sequence
Tā shuō <u>fǎyī</u> mǎshàng jiù chūfā.
He says the forensic doctor will set out immediately.
(2) LhL sequence (h = weakened H)
Nǐ ràng tā <u>mǎyīmǎ</u> zhèduī zhuāntou.
Tell him to stack up this pile of bricks.
(3) LRL sequence
Tā xiǎng <u>zhuǎnyí</u> nǐ de shìxiàn.
He is trying to divert your attention.

Table 1. Sample stimulus sentences. The underscored words are to be emphasized by the subjects during recording. The tone marks “ ˇ , ˉ , ˊ and ˋ ” represent H, R, L and F, respectively. Syllables with the neutral tone have no tone marks.

2.1.2. Subjects. Four native speakers of Beijing Mandarin, two males and two females, participated as subjects. They were graduate students at Northwestern University. All of them were in their twenties.

2.1.3. Procedures. Recording was conducted in a sound-treated booth in the Speech Acoustics Laboratory at the Department of Communication Sciences and Disorders, Northwestern University. The target sentences were presented in Chinese in random order on a Macintosh Performa 6400/180 using a custom-written program. The computer monitor and the microphone were placed in a sound-treated booth, and the subject was seated in front of the monitor. The target word/phrase in the sentence was underscored as shown in Table 1, indicating where the subject should put the focus. The subject read the target sentence aloud 3 times in each trial, first at a normal speaking rate, then at a fast rate, and finally at a slow rate. The subject was instructed not to pause in the middle of a sentence even at the slow speaking rate. The sentences were directly digitized into a Macintosh 7500/100 at a sampling rate of 22 kHz.

2.3. Analysis and Results

The effect of tone and speaking rate on syllable duration was first examined. Because the boundary between syllables 1 and 2 is mostly ambiguous due to the initial semivowel /y/ in syllable 2 [12], it was difficult to measure the duration of syllable 2 (which carried the target tone) separately from that of syllable 1. So, instead, the combined duration of syllables 1 and 2 was measured and compared for each subject. It was found that the duration values were well separated across the three speaking rates. Two-factor ANOVAs with duration as dependent variable and tone and speaking rate as independent variables were also conducted. Both effects were significant at $p < .0001$ for all subjects, indicating that the recording procedure effectively elicited systematic duration variations from the subjects.

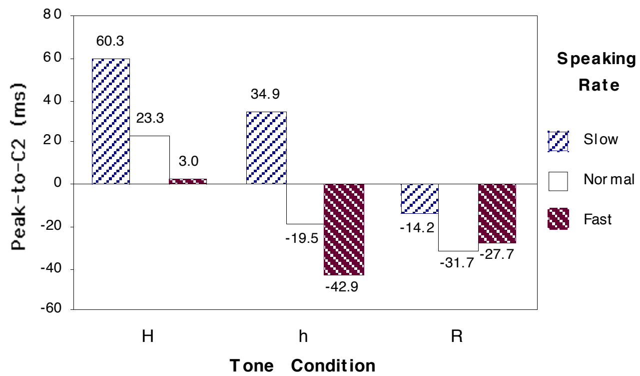

Figure 1. Mean Peak-to-C2 values (averaged across subjects) in different tone and speaking rate conditions.

2.3.1. Magnitude and frequency of occurrence of peak delay. The effects of tone and speaking rate on peak delay were then examined. F_0 tracings of the recorded sentences were obtained using a procedure described in [14, 15]. Peak delay was indicated by Peak-to-C2, which measures the location of F_0 peak relative to the onset of the initial sonorant in syllable 2 (C2). A negative value of Peak-to-C2 indicates peak delay. Figure 1 displays Peak-to-C2 in different tone and speaking rate conditions averaged across all subjects. As can be seen in Figure 1, across all speaking rates, the mean peak location is later than the onset of C2 in R but earlier than C2 in H. The mean peak location in h, however, depends much on speaking rate: extensively delayed at normal and fast speaking rates, but well within syllable 2 at slow speaking rate.

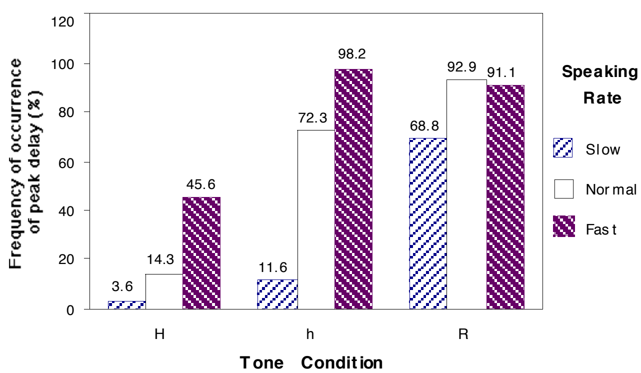

Figure 2. Frequency distribution of peak delay (averaged across subjects) in different tone and speaking rate conditions.

That the mean Peak-to-C2 values in H are positive at all speaking rates does not mean that peak delay never occurred in H. In fact it occurred quite often at the fast speaking rate. This is seen in the frequency of occurrence of peak delay, which is obtained by counting the number of instances of negative Peak-to-C2 in each condition. Figure 2 displays the mean frequency of occurrence of peak delay in all conditions (averaged across subjects). As can be seen in Figure 2, peak delay is not an all-or-none event. Rather, it occurred more or less in all tone and speaking rate conditions. At the same time, however, peak delay occurred much more frequently in R and h than in H, and at faster speaking rates than at slower speaking rates.

2.3.2. Peak alignment. To examine how closely peak delay is related to the sharp rise near the end of a syllable, a set of linear regression analyses were conducted. First, Peak-to-C2 was regressed on Max-velocity-to-C2 — the distance between the point of maximum velocity in the F_0 curve and the onset of syllable 3. The point of maximum velocity in an F_0 curve was located by taking the first derivative of the curve. The regression results for all subjects are shown in Table 2. As can be seen, the r^2 values are high in general (except for those of R for subjects ZY and LX). The high overall r^2 values indicate that much of the variation in peak location can be related to the location of maximum rising velocity relative to the offset of syllable 2: the later the velocity peak, the later the F_0 peak.

Tone condition	Subject	Coefficient	Intercept	r^2
H	DX	.48	-33.52	.66
	LX	.37	-22.12	.71
	YJ	.36	-18.71	.71
	ZY	.27	-17.63	.31
h	DX	.96	-87.95	.61
	LX	.72	-62.69	.90
	YJ	.79	-60.09	.94
	ZY	.71	-75.11	.82
R	DX	.78	-62.31	.49
	LX	.30	-35.37	.28
	YJ	.44	-49.61	.40
	ZY	.37	-50.69	.16

Table 2. Results of linearly regressing Peak-to-C2 on Velocity-peak-to-C2.

If the occurrence of peak delay is closely related to the location of a sharp F_0 rise near the end of a syllable, it may be possible to assess the critical distance between the rising contour and the syllable offset at which peak delay may occur. Besides Velocity-peak-to-C2 as shown in Table 2, the rise location can be also indicated by the onset of the rise. The onset of the rise can be defined as the point at which the F_0 contour takes a sharp turn upwards. Mathematically, that point corresponds to the location of maximum acceleration in the F_0 curve and can be located by taking the second derivative of the curve. Linear regression equations were thus obtained using either Max-velocity-to-C2 or Max-acceleration-to-C2 as the predictor and Peak-to-C2 as the dependent variable. The critical distances were then computed by solving the equations for the values of the predictors at Peak-to-C2 = 0. The coefficients and intercepts in the regression of Max-acceleration-to-C2 on Peak-to-C2 are shown in Table 3. Table 4 displays the estimated critical distances computed using the parameters in Tables 2 and 3.

In most cases in Table 4, the critical distances for H are smaller than h, which are in turn smaller than R. Some of the critical distances for R are unreasonably large (in particular, 680.6 ms for ZY), given that peak delay actually occurred regularly in R (as shown in Figures 1 and 2). Those large values are mostly due to the shallow slope (i.e. small coefficient) of the respective regression lines. Apparently, these values should not be considered as the true critical distances for R. In fact, for most subjects, Max-acceleration-to-C2 and Max-velocity-to-C2 in R generally remain close to the critical values estimated from H and h. Across all subjects, the largest mean Max-acceleration-to-C2 is 128.3 ms for R, while the smallest mean critical Max-acceleration-to-C2 is 125.7 ms in Table 4. The largest mean Max-velocity-to-C2 in R is 70.5 ms, while the smallest mean critical Max-velocity-to-C2 is 61.7 ms in Table 4. It therefore seems that, given the smallest mean critical distances in Table 4, peak delay in fact *ought to* regularly occur in R: because its final F_0 rise consistently occurred near or within the critical distances from the syllable offset.

Tone condition	Subject	Coefficient	Intercept	r^2
H	DX	0.35	-40.89	0.60
	LX	0.41	-61.29	0.73
	YJ	0.33	-32.51	0.66
	ZY	0.31	-42.13	0.39
h	DX	0.89	-127.99	0.58
	LX	0.54	-76.79	0.73
	YJ	0.67	-89.13	0.75
	ZY	0.78	-126.87	0.90
R	DX	0.31	-53.07	0.27
	LX	0.09	-26.81	0.05
	YJ	0.42	-66.88	0.44
	ZY	0.07	-45.60	0.04

Table 3. Results of linearly regressing Peak-to-C2 on Acceleration-peak-to-C2.

Estimated Critical value		Max-acceleration-to-C2			Max-velocity-to-C2		
Tone Condition		H	h	R	H	h	R
Subject	DX	117.2	143.3	170.7	69.8	92.0	79.7
	LX	148.8	141.2	308.2	59.8	86.6	116.3
	YJ	98.5	132.6	161.2	52.0	76.5	113.8
	ZY	138.1	161.8	680.6	65.3	105.9	137.4
	Mean	125.7	144.7	330.2	61.7	90.3	111.8

Table 4. Critical values of Max-acceleration-to-C2 and Max-velocity-to-C2 for each tone produced by each subject.

3. DISCUSSION

The results of peak delay analyses revealed interesting information about peak delay in Mandarin. The magnitude as well as frequency of occurrence of peak delay was found to vary across both tone and speaking rate conditions. Alignment analyses revealed that peak delay was closely related to the location of the final F_0 rise in a syllable. Peak delay was more likely to occur if the F_0 rise was close to the offset of the tone-carrying syllable. An assessment was also made of the critical distance between the F_0 rise and the syllable offset at which peak delay may occur. Regardless of the target tone, peak delay was likely to occur if the F_0 rise started at about 125 ms or closer from the end of the syllable, or if the rising velocity peaked at about 60 ms or closer from the syllable offset.

These findings seem to support the hypothesis that peak delay in Mandarin can be mostly attributed to the interaction between tonal targets and their articulatory implementation. While there are overall differences in terms of peak delay among the tone conditions in this study, peak delay in each tone also varies with speaking rate. The patterns of peak delay variation seem to be accountable in terms of simple articulatory constraints on the implementation of the tonal targets. If a combination of the target tone and its tonal context necessitates a sharp rise in F_0 and if that rise is forced to occur too close to the syllable offset, peak delay is likely to occur, presumably because the larynx cannot terminate a sharp F_0 rise instantaneously.

The findings of the present study seem to further support a more general hypothesis: that F_0 peak alignment in Mandarin in general can be accounted for by the interaction between underlying pitch targets and their articulatory implementation [16]. As argued in [15, 16], the acoustic manifestation of a tone may not always directly reflect its underlying pitch target. The observed F_0 patterns are probably the results of the speaker's implementation of the underlying tonal targets within the time frame of the segmental units designated to carry them. In the case of R, the underlying pitch target is rising, which is implemented the fullest by the end of the R-carrying syllable. When R is followed by L, the implementation of L (which has a low pitch target) starts as soon as the R-carrying syllable ends. However, because it takes time for the larynx to terminate the sharp rise, the actual turning point occurs somewhat *after* the R-L syllable boundary. Thus in a RL sequence, the consistent peak delay is likely due to the consistent alignment of the two tones with their respective host syllables. The present study finds that the F_0 peak in H can also be pushed out of the syllable at fast speaking rate. This further confirms that articulatory constraints play an important role in determining the shape and alignment of F_0 contours.

To summarize the above discussion, schematic peak alignments in F, H and R in Mandarin are displayed in Figure 3. As shown in the upper panel of Figure 3, any tone with a high pitch in its underlying target has a potential of generating an F_0 peak, given the right tonal context. But the relative location of the peak depends on the inherent properties of the pitch target. In F the peak is relatively early because it has a falling target, whereas the peak in R is relatively late due to its rising target. For H, its F_0 contour tends to level off late in the syllable due to its static high target. However, as shown in the lower panel of Figure 3, at fast speaking rate, the rising slope in H may be pushed close to the end of the syllable. As a result, the peak in H may occur either very close to, or even somewhat after the syllable boundary at fast speaking rate.

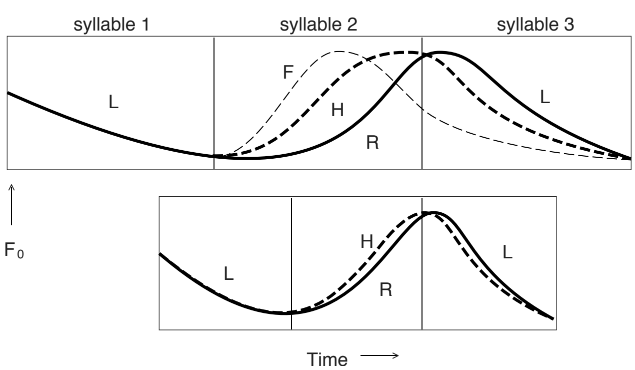

Figure 3. Schematic peak alignments in F, H and R at normal speaking rate (upper panel), and in H and R at fast speaking rate (lower panel).

4. CONCLUSION

Previous studies have found evidence that the occurrence and general location of F_0 peaks in Mandarin are explainable in terms of the interaction between underlying tonal targets and articulatory constraints [13, 14, 15]. The findings of the present study provide evidence that peak delay — a finer peak alignment pattern — is also explainable in terms of such interaction. Certain underlying pitch target and tonal context may necessitate a sharp rising contour near the end of a syllable. If such a rise is forced to occur too close to the syllable offset, either by the requirement of underlying target (in case of R), or by shortened syllable duration (at fast speaking rate or when the syllable is prosodically weak), peak delay may result. Hence, an articulatory constraint — that it takes time for the larynx to reverse the direction of F_0 movement — may interacts with tonal targets and tonal context to cause peak delay to occur in rather different conditions.

NOTES

[1] 'Peak delay' is sometimes used in the literature to refer to a measurement that indicate the location of an F_0 peak relative to the onset of a syllable or rhyme. That definition is not used in the present paper.

[2] The tone of the middle syllable may even be considered as neutralized. The neutral tone in Mandarin is said to have no definite pitch target of its own, and its F_0 value varies depending on the tone of the preceding syllable [2]. After a L tone, the neutral tone has a high F_0 value. As will be seen, the height of the F_0 peak in the weakened H tone and in the regular H tone was not very different.

REFERENCES

[1] Arvaniti, A., Ladd, D. R. and Mennen, I. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, 36, 3-25.

[2] Chao, Y. R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

[3] Gårding, E. 1977. The importance of turning points. In L. M. Hyman (ed.), *Studies in Stress and Accent*. Los Angeles: Department of Linguistics, University of Southern California. 27-35.

[4] Gårding, E. and Zhang, J. 1986. Tone 4 and Tone 3 discrimination in Modern Standard Chinese. *Language and Speech*, 29, 281-293.

[5] Gårding, E. 1987. Speech act and tonal pattern in Standard Chinese. *Phonetica*, 44, 13-29.

[6] Grimm, C. 1997. Pitch accent in Oneida. Presentation at the 1997 Annual Meeting of the Linguistic Society of America, Chicago.

[7] de Jong, K. 1994. Initial tones and prominence in Seoul Korean. *Ohio State University Working Papers in Linguistics*, 43, 1-14.

[8] Kim, S.-A. 1999. Positional effect on tonal alternation in Chichewa: Phonological rule vs. phonetic timing. In *Proceedings of Chicago Linguistic Society*, Vol. 34.

[9] Liberman, M. and Pierrehumbert, J. 1984. Intonational invariance under changes in pitch range and length. In Aronoff, M. and Oehrle, R. (Eds.), *Language Sound Structure*. Cambridge, MA: MIT. Press. 157-233.

[10] Prieto, P., Santen, J. v. and Hirschbert, J. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23, 429-451.

[11] Silverman, K. E. A. and Pierrehumbert, J. B. 1990. The timing of prenuclear high accents in English. In Kingston, J. and Beckman, M. E. (eds.), *Papers in Laboratory Phonology 1 — Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press. 72-106.

[12] Xu, Y. 1986. Putonghua yinlian de shengxue yuyinxue texing (Acoustic-phonetic characteristics of junctures in Mandarin). *Zhongguo Yuwen*, 353-360.

[13] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61-83.

[14] Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, 55, 179-203.

[15] Xu, Y. in press. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*.

[16] Xu, Y. and Wang, Q. E. in press. Pitch targets and their realization: Evidence from Mandarin Chinese. Speech Communication.