Original Paper

Phonetica 2001;58:26–52

Received: March 5, 1999
Accepted: March 7, 2000

Fundamental Frequency Peak Delay in Mandarin

Yi Xu

Department of Communication Sciences and Disorders, Program of Speech and Language Pathology, Northwestern University, Evanston, Ill., USA

Abstract

Fundamental frequency (F_0) peak delay (henceforth peak delay) refers to the phenomenon that an F_0 peak sometimes occurs after the syllable it is associated with either lexically or prosodically. Although peak delay has been reported for various languages, the mechanism of its occurrence has so far remained unclear. In Mandarin, peak delay has been found to occur regularly in the rising (R) tone but not in the high (H) tone. The present study investigates the underlying mechanisms of peak delay by examining its relationship with tone, tonal context, and speech rate. An experiment was conducted to test the possibility that peak delay may occur also in the H tone in Mandarin if the H-carrying syllable is sufficiently shortened. Native speakers of Mandarin Chinese recorded three types of sentences each containing a H, a R, or a weakened H tone. The sentences were produced at three speech rates: normal, fast and slow. Analysis of F_0 contours and peak alignment revealed that at normal speech rate, peak delay occurred regularly in both the R and weakened H tones but only occasionally in the H tone; at slow speech rate, peak delay continued to occur regularly in the R tone but only occasionally in weakened H and rarely in the H tone; at fast speech rate, peak delay occurred not only regularly in the R and weakened H tones, but also frequently in the H tone. Results of F_0 contour alignment analysis indicate that peak delay occurs when there is a sharp F_0 rise near the end of a syllable, regardless of the cause of the rise. The finding of this study provides support for the view that, rather than due to actual misalignment between underlying pitch units and segmental units, much of the variability in the shape and alignment of F_0 contours in Mandarin is attributable to the interaction of underlying pitch targets with tonal contexts and articulatory constraints.

Introduction

A tone, a pitch accent or a focal prominence may often generate a peak in the fundamental frequency (F_0) tracing of an utterance, as illustrated in figure 1. If the tone, pitch accent, or focal prominence is carried by a syllable, by the simplest assumption,

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/01/0582–0026 $17.50/0
Accessible online at:
www.karger.com/journals/pho

Yi Xu, 2299 N. Campus Drive
Communication Sciences and Disorders
Northwestern University, Evanston, IL 60208 (USA)
Tel. +1 (847) 491-2416, Fax +1 (847) 467-2776
E-Mail xuyi@northwestern.edu

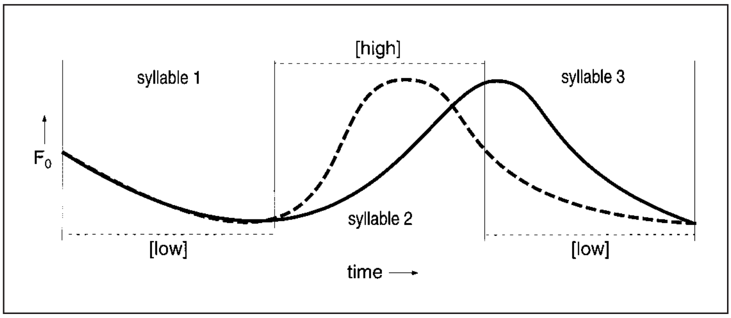


Fig. 1. A schematic illustration of F_0 peak delay. The [high] and [low] are the underlying pitch targets, and the curved lines are the surface realization of the pitch targets. The F_0 peak in the solid line is delayed beyound the end of the [high]-carrying syllable, while the F_0 peak in the dashed line is not. (The F_0 curves in the graph are meant to be conceptual only, and thus are not necessarily consistent in all aspects with what is already known from previous studies.)

the F_0 peak associated with it should occur somewhere inside that syllable, as illustrated by the dashed curve in figure 1. However, it has been noted in past studies that the peak often occurs after the syllable that carries the tone, pitch accent or focal prominence, as illustrated by the solid curve in figure 1. The present paper is concerned with this phenomenon and will refer to it as 'peak delay'.[1]

Peak delay has been reported, explicitly or implicitly, for a number of languages. Silverman and Pierrehumbert [1990] found that an F_0 peak associated with a prenuclear H* pitch accent in English often occurred after the syllable that carries the pitch accent; de Jong [1994] found that an F_0 peak associated with the initial tone in Korean often occurred after the end of the initial syllable. Prieto et al. [1995] reported that in Mexican Spanish an F_0 peak associated with a H* accent often occurred after the stressed syllable unless the syllable was immediately followed by another stressed syllable in the following word. Grimm [1997] found that in Oneida, an F_0 peak associated with a pitch accent usually occurred in the syllable following the stressed syllable if the vowel in the stressed syllable was short. Arvaniti et al. [1998, p. 23] reported that in Greek 'the H target is very precisely aligned just after the beginning of the first postaccentual vowel'. Kim [1999] found that in Chichewa F_0 peaks occurred consistently right after the offset of the H-bearing syllable if the syllables were prepenult.

The main concern of most of the aforementioned studies, as well as some earlier ones on F_0 peaks, however, is not peak delay (as defined above) per se, but the alignment of F_0 peaks with segmental units in general. Steele [1986] noted that in English the location of a peak associated with H* relative to the onset of the final stressed vowel (i.e., with nuclear-stress) in an utterance varied with the duration of the vowel, but the patterns of variation were different depending on the cause of the vowel-duration change. When the duration was shortened by faster speech rate or by intrinsically

[1] 'Peak delay' is sometimes used in the literature to refer to a measurement that indicates 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.

shorter vowels, the peak became closer to the onset of the stressed vowel. When the vowel duration was shortened by added postnuclear syllables, the F_0 peak either moved away or maintained same distance from the vowel onset. Silverman and Pierrehumbert [1990] found similar patterns for nonfinal stressed vowels (with prenuclear stress). Prieto et al. [1995] found comparable peak alignment patterns in Mexican Spanish. That is, in general, as the duration of vowel and consonant in a stressed syllable increased, the peak moved further away from the syllable onset. However, when the vowel was lengthened by an upcoming word or phrasal boundary, the peak actually shifted closer to the syllable onset. Similarly, F_0 peaks were found to occur consistently right after the H-carrying syllable in Chichewa only when the tone was prepenult but not when it was penult [Kim, 1999]. In Oneida, peak delay occurred in phonological short vowels but not long vowels [Grimm, 1997].

It therefore seems that there is a possible dichotomy between peak alignment variations due to phonetic factors, such as speech rate and intrinsic vowel duration, and those due to prosodic factors, such as syllable position, word and phrase boundaries and other prosodic contexts. In general, when a syllable is lengthened due to phonetic factors, such as speech rate or intrinsic vowel length, the F_0 peak moves away from the syllable or vowel onset along with the syllable offset. In contrast, when a syllable is lengthened due to prosodic factors, such as an upcoming stress, pitch accent or phrase boundary (or the lack thereof), or due to *phonological* vowel lengthening, the F_0 peak moves in the other direction: closer to the syllable onset than to the offset.

This kind of dichotomy appears to parallel certain peak alignment variations in Mandarin tones. Xu [1998, 1999] found that peak alignment variations due to tone and those due to speech rate and focal prominence in Mandarin were quite different. On the one hand, when a syllable was lengthened or shortened due to changes in speech rate or focal prominence, the F_0 peak in the high (H), rising (R) and falling (F) tones all moved with the syllable offset rather than with the onset. On the other hand, peak alignments of the three tones had very different overall patterns. In general, the peak in the F tone occurred near the middle of the syllable; the peak in the H tone occurred close to but *before* the end of the syllable, and the peak in the R tone occurred immediately *after* the end of the syllable (i.e. delayed). This parallel between Mandarin and other languages could be just superficial. But it is also possible that it is due to certain basic mechanisms that underlie the generation and use of pitch contours in speech in general.

To understand these mechanisms, Xu and Wang [in press] proposed that (a) underlying the phonemic lexical tones are articulatorily operable units comparable to segmental phones, (b) these underlying pitch targets are either static (such as [high] or [low]) or dynamic (such as [rise] or [fall]), (c) these pitch targets are synchronously implemented with the syllables that carry the lexical tones, and (d) the manner of pitch-target implementation is such that the underlying targets are always asymptotically approximated over time. Based on these proposals, the F and R tones would naturally have diametrically different peak alignments because the pitch target [fall] in the F tone necessitates a relatively early high point in a syllable while the target [rise] in the R tone necessitates a relatively late high point. Likewise, peak alignment in the H tone is later than in the F tone but earlier than in the R tone because of the target [high] does not have to be either extremely early or late. In other words, the three distinct peak alignment patterns of the F, R and H tones may be directly attributable to the different properties of the pitch targets underlying these tones. Also based on these proposals, the synchronous implementation of tones with their carrying syllables would naturally lead to a positive correlation between the location of F_0 peaks and syllable duration: the longer the syllable, the later the peak relative to the syllable onset, as has been found by Xu [1998, 1999].[2] Furthermore, the proposed synchronous implementation of tone with the syllable and asymptotic approximation of pitch targets would account for the general pattern of contextual tonal variations reported in Xu [1997], i.e., contextual tonal variations in Mandarin are mostly preservative, and much of the F_0 contour of a syllable constitutes a continuous transition from the offset F_0 of the preceding syllable to a contour that is appropriate for the tone of the current syllable.

Xu and Wang [in press] further attempted to account for variations in the occurrence of peak delay in Mandarin. As mentioned earlier, peak delay has been found to occur regularly in the R tone in Mandarin, but not in the H tone. Close examination of F_0 contours revealed that in the R tone the main portion of the rise always occurred in the final portion of the syllable, regardless of syllable duration [Xu, 1998, 1999]. In contrast, the F_0 contour in the H tone (when preceded by the L or F tone) tends to level off toward the end of the syllable [Xu, 1999]. Articulatorily, if a sharp F_0 rise is implemented in the final portion of a syllable in the R tone, and when the following tone is L, a sharp turn needs to be made at the syllable boundary. To produce such a sharp turn, the larynx needs to first stop the pitch raising gesture and then start a pitch lowering gesture. This process should take time, and the result would be that the peak (i.e., the turning point) actually occurs in the following (L-carrying) syllable. In contrast, peak delay does not occur in the H tone probably because there is enough time for the transition from the preceding [low] (in the L tone) to the current [high] to be completed so that the rising F_0 contour in the H tone can level off before the end of the syllable. Thus the sharp F_0 rise just before the syllable offset may play a critical role in causing the peak delay in the R tone, and the lack of such sharp F_0 rise just before the syllable offset probably prevents peak delay from occurring in the H tone. Here again, the interaction of phonological and articulatory factors probably determines the occurrence of peak delay: the underlying target and the tonal context stipulate where a sharp rise should occur in a syllable, while the articulatory implementation determines whether or not peak delay actually occurs.

A natural implication 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, if the H-carrying syllable in a LHL sequence is substantially shortened for some reason, the F_0 rise may be pushed close to the end of the syllable just as in the R tone in a LRL sequence. In such a case, would peak delay also occur in the H tone, just as in the R tone? Answering this question may help verify the interactive account of F_0 peak alignment in Mandarin and possibly in other languages as well. The following experiment was therefore designed to investigate how peak delay relates to tone, tonal context, and syllable duration.

[2] Xu [1998] also found that the dynamic targets in the R and F tones are always fully implemented in the later portion of a syllable regardless of syllable duration. As syllable duration gets longer due to reduced speech rate, speakers delay the onset of rise in the R tone, as if to guarantee that the most dynamic portion of the F_0 contour occurs by the end of the syllable. Nevertheless, even at slow speech rate, F_0 *starts* its initial drop in the R tone in an HR sequence immediately after the syllable boundary, indicating that the synchrony between tone and syllable is maintained regardless of speech rate.

Table 1. Sentences recorded in the experiment

(1) LHL sequence	
Tā shuō fǎyī mǎshàng jiù chūfā.	He says the forensic doctor will set out immediately.
Wǒmen nàge zhōu mǎi yī miǎn shuì.	Buying clothing in my state is tax-free.
Gàosu wǒ tǒngyī Měiguó shì nǎ yī nián.	Tell me when America was united.
Bùdǒng yīlǐ bùnéng kànbìng	One can't treat patients without knowing medical principles.
(2) LhL sequence (h = weakened H)	
Nǐ ràng tā mǎ yī mǎ zhèduī zhuāntou.	Tell him to stack these bricks.
Zhè yào děi niǎn yī niǎn cái kéyǐ yòng.	This medicine must be ground before use.
Shǐjìng niǔ yī niǔ zhège ménbǎ.	Turn the doorknob firmly.
Bǎ nǐ de jiǎo nuǎn yī nuǎn zài chū mén.	Warm up your feet before you leave.
(3) LRL sequence	
Zhè zhuāng kěyí mǎimai yīdìng yào cháqīng.	This suspicious business deal must be investigated.
Wǒxiǎng zhèxiē lǐyí miǎn bù liǎo.	I think these formalities cannot be dispensed with.
Nǐ de xiǎiyímǔ jīntiān xiàwǔlái.	Your aunt will come this afternoon.
Tā xiǎng zhuǎnyí nǐ de shìxiàn.	He is trying to divert your attention.

These sentences are divided into three groups. Group 1 contains a LHL sequence; group 2 contains a LhL sequence, where h denotes a weakened H tone, and group 3 contains a LRL sequence. The underscored words were where subjects put the focus while recording the sentences. The transcription is in Pinyin, the official spelling system for Mandarin pronunciation.

Number of utterances: 3 (tone sequence) × 4 (sentences) × 3 (speech rates) × 7 (repetitions) = 252 (utterances).

The Experiment

Method

Stimuli

The basic design of the experiment was to manipulate target tone, tonal context and syllable duration to see how F_0 peak alignment is affected. Three target tones were used: R, H and weakened H. A weakened H occurs in Mandarin in the middle syllable of a verb phrase such as 'nuǎn yī nuǎn' 'to warm up a bit'. The middle syllable in this kind of phrase structure is prosodically weak and usually has shortened duration.[3] Tonal context was controlled by using a L_L frame held constant across all three tone conditions. This contextual frame ensures the occurrence of an F_0 peak in or near the middle syllable in every utterance [Xu, 1999]. Syllable duration was controlled mainly by changing speech rate. Three speech rates were used, normal, fast and slow. In addition, as just mentioned, the weakened H was also expected to shorten syllable duration. Table 1 displays all the sentences used in the experiment. These sentences are divided into three groups. Group 1 contains a LHL sequence; group 2 contains a LhL sequence, where h denotes a weakened H tone, and group 3 contains a LRL sequence.[4] The underscored words in table 1 were to be focused by the subjects while recording the sentences. The purpose of specifying the focus was to guarantee that there would always be a clear peak in the F_0 contour near the target tone. In order not to confuse the subjects, only whole words were underscored. As can be seen, the syllable carrying the target tone (H, R, or weakened H) is always

[3] 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 [Chao, 1968]. After L, the neutral tone has a high F_0 value. As will be seen, the height of the F_0 peak in the weakened H and in the regular H was not very different.

[4] Although the L tone in Mandarin is realized as having a final rise after the low F_0 point when said in isolation and sometimes in prepausal positions, no final rise occurs when the L tone is said before other tones.

under focus, while the second L tone in a sequence is either on-focus or postfocus. The effect of focus is to increase the pitch range of the focused word and suppress (compress and lower) the pitch range of the postfocus words [Garding, 1987; Jin, 1996; Xu, 1999]. So, for all the target tones, the F_0 maximum would be raised by the focus. For the L tone following the target tone, the F_0 minimum would be lowered whether it is on-focus or postfocus, since both pitch range expansion and suppression would lower F_0 for the L tone. As indicated in table 1, the sentences were repeated seven times in blocks, each with a different randomization order.

Because of the difficulty in finding words with the desired CV structure and tone, test sequences having different syntactic structures had to be used. According to Shih [1986] and Speer et al. [1989], in Mandarin, the prosodic closeness of adjacent tones is predominantly determined by the prosodic structure of a sentence rather than the syntactic structure. Furthermore, Shih and Sproat [1992] also found that the prosodic structure also determines the strength of each syllable.

Most of the target words/phrases in the LHL and LRL sequences are in similar prosodic structures, where the target syllable /yi/ occurs at the right edge of a disyllabic prosodic foot. There are two exceptions. The first is in the last sentence in the LHL sequence in table 1, where the /yi/ is at the left edge of a prosodic foot. The second exception is in the third sentence in the LRL sequence, where the /yi/ is the middle syllable in a trisyllabic foot. While it was due to difficulty in finding appropriate target words/phrases that these two sentences were used, it would be interesting to see if the prosodic variations made any difference in the alignment of the F_0 peaks.

Subjects

Four native speakers of Beijing Mandarin, 2 males and 2 females, participated as subjects. They were all graduate students studying at Northwestern University. All of them had lived in Beijing, China, since their childhood, and they were all in their twenties.

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 speech signal was received by a condenser microphone connected to a Macintosh 7500/100 computer. The signal was digitized at 22 kHz in real time using the SoundEdit program (Macromedia, Inc.) and stored on the computer’s hard disk.

The target sentences were presented in Chinese on a Macintosh Performa 6400/180 computer using a custom-written Java program. The computer monitor was placed in a sound-treated booth. The subject was seated in front of the monitor. The microphone was placed by the side of the monitor, approximately 30 cm away from the subject’s lips. At the start of the session, the subject clicked the ‘Start’ button on the monitor screen with a mouse to display a set of instructions on the screen. After reading the instructions, the subject did a number of practice trials to get familiarized with the procedure and requirements. In each trial, the subject clicked an on-screen button to bring up a sentence in Chinese on the screen. 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 speech rate, then at a fast rate, and finally at a slow rate. (This method was found to be able to elicit utterances with clear durational differences [Xu, 1998].) The subject was instructed not to pause in the middle of a sentence even at the slow speech rate. Since all the sentences were quite short, subjects had little difficulty following this instruction. In case a mistake was made, the subject was asked to repeat the same sentence at all three speech rates. The sentences were presented in random order, and the order of stimulus presentation was different for each subject.

F_0 Extraction and Various Measurements

The digitized signals were transferred from the Macintosh computer to a Sparc 5 workstation. The signal was converted to a format readable by programs in the ESPS/waves + signal processing software package (Entropic Inc.). The individual target sentences were extracted and saved as separate ESPS signal files. The ESPS *epochs* program was then used to mark every pitch period in the target words. After that, the marked signals were manually edited using the ESPS *xlabel* program attached to the ESPS *xwaves* program to correct spurious vocal pulse markings such as double-marking or pitch-period skipping.

While checking for spurious vocal pulse markings, segmentation labels were also added to mark the boundaries between certain segments in the target word. The boundaries marked were the onset of syllable 1, the vowel onset in syllable 1, the onset of syllable 3, the vowel onset in syllable 3, and the offset of syllable 3. Since the boundary between the syllable /yi/ and the preceding syllable is ambiguous in most cases in Mandarin [Xu, 1986], the boundary between the first and second syllables was not marked.

The vocal pulse markings and segment labels for each utterance were saved in a text file. Those files were then processed by a set of custom-written C programs. The programs first converted the duration of pitch periods into F_0 values, and then smoothed the resulting F_0 curves using a *trimming algorithm* [Xu, 1999] that eliminates sharp bumps and edges. The smoothed F_0 curves were then subjected to further analysis.

Another C program was written to take several measurements to be used in statistical analysis. The first measurement is syllable duration. Because of the aforementioned ambiguity in syllable boundary between the first and second syllables, only the combined duration of syllables 1 and 2 was measured. The second measurement is Peak-to-C3, which is the location of the F_0 peak relative to the onset of the initial consonant in syllable 3 (C3). The value is positive if the F_0 peak occurs earlier than C3, negative if later than C3. The F_0 peak was located by finding the maximum value in the smoothed F_0 curve. The third measurement is Max-velocity-to-C3, the distance between the point of maximum velocity in the F_0 curve and the onset of the initial consonant in syllable 3. The point of maximum velocity was found by taking the first derivative of each F_0 curve. The fourth measurement is Max-acceleration-to-C3, the distance between the point of maximum acceleration in the F_0 curve and the onset of the initial consonant in syllable 3. The point of maximum acceleration was found by taking the second derivative of each F_0 curve.

Analysis and Results

Comparison of F_0 Curves

To graphically examine F_0-syllable alignment, the smoothed F_0 curves of the three syllables were averaged across the seven repetitions spoken at the same speech rate using the following procedure. First, for each segment, the mean duration was computed across the seven repetitions. Next, a normalized mean F_0 contour of each segment was computed by taking a fixed number of points at equal time proportions and averaging them across the seven repetitions. The mean F_0 curves of all segments were then plotted as a function of real time and aligned by different segment locations for visual inspection. Because the only thing averaged out when obtaining these plots was the variability in F_0 and duration across the seven repetitions, everything that was consistently produced across the repetitions was fully preserved in these plots. Figure 2 displays the mean F_0 contours from all 4 subjects. Among them, S1 and S2 are the male subjects, and S3 and S4 are the female subjects. These F_0 contours are aligned by the onset of syllable 1, and the gap between time 0 and the beginning of each curve corresponds to the duration of the initial voiceless consonant of syllable 1. The dotted region in each curve corresponds to the initial sonorant of syllable 3. The portion of the F_0 curve before the dotted region corresponds to the rhyme of syllable 1 and the entire syllable 2, and the portion after the dotted region corresponds to the rhyme of syllable 3.

Several observations can be made of the F_0 contours in figure 2. The first is that, as expected, for most of the sentences, there is usually a sharp F_0 drop after the onset of the F_0 contour. Presumably, partly due to the F_0-raising effect of the initial voiceless consonants [Lehiste and Peterson, 1961; Hombert, 1978; Santen and Hirschberg, 1994] and partly due to the nonlow ending of the preceding tone [Xu, 1994, 1997], the onset F_0 in syllable 1 was not low enough for the L tone. The F_0 after the syllable onset there-

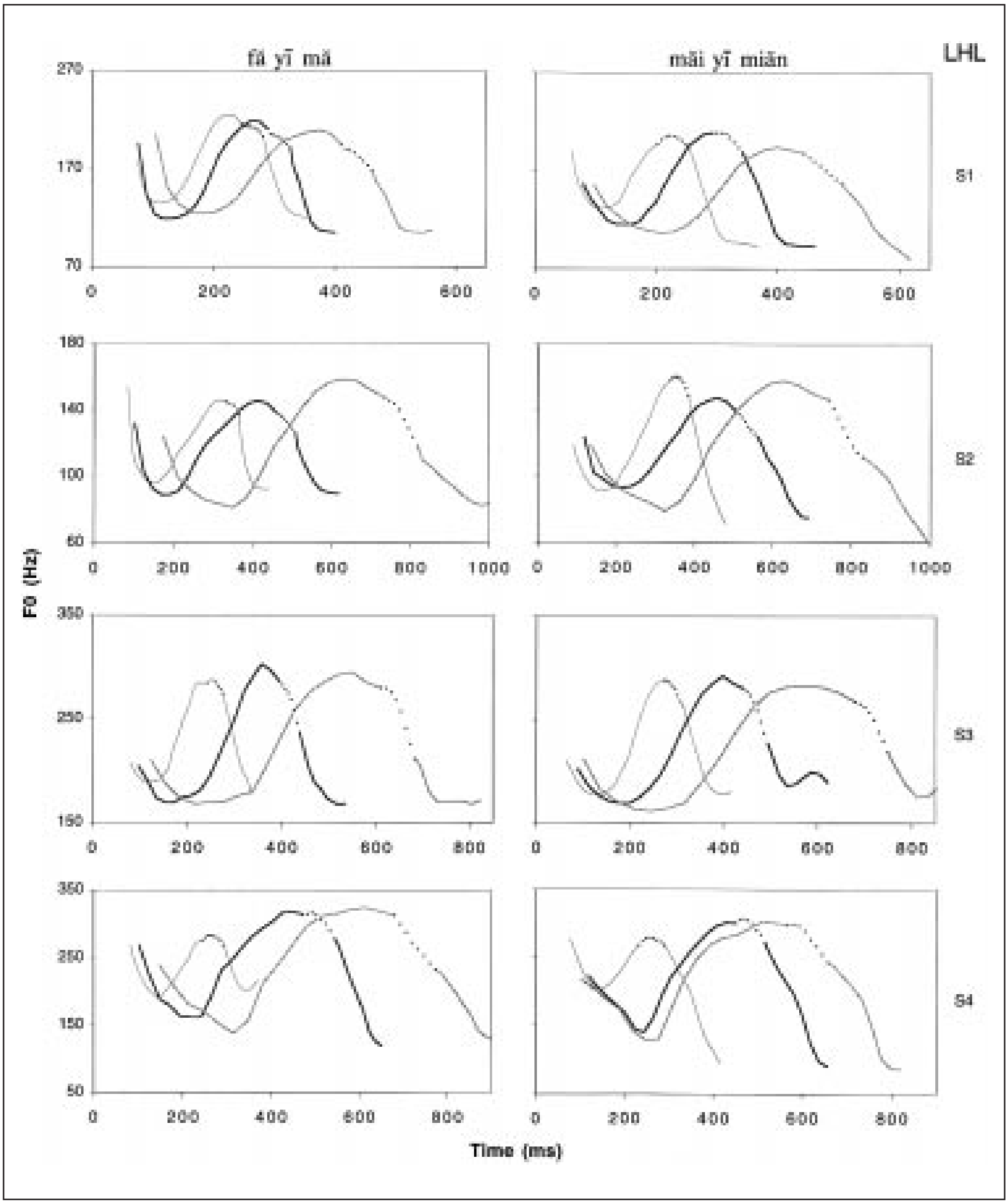
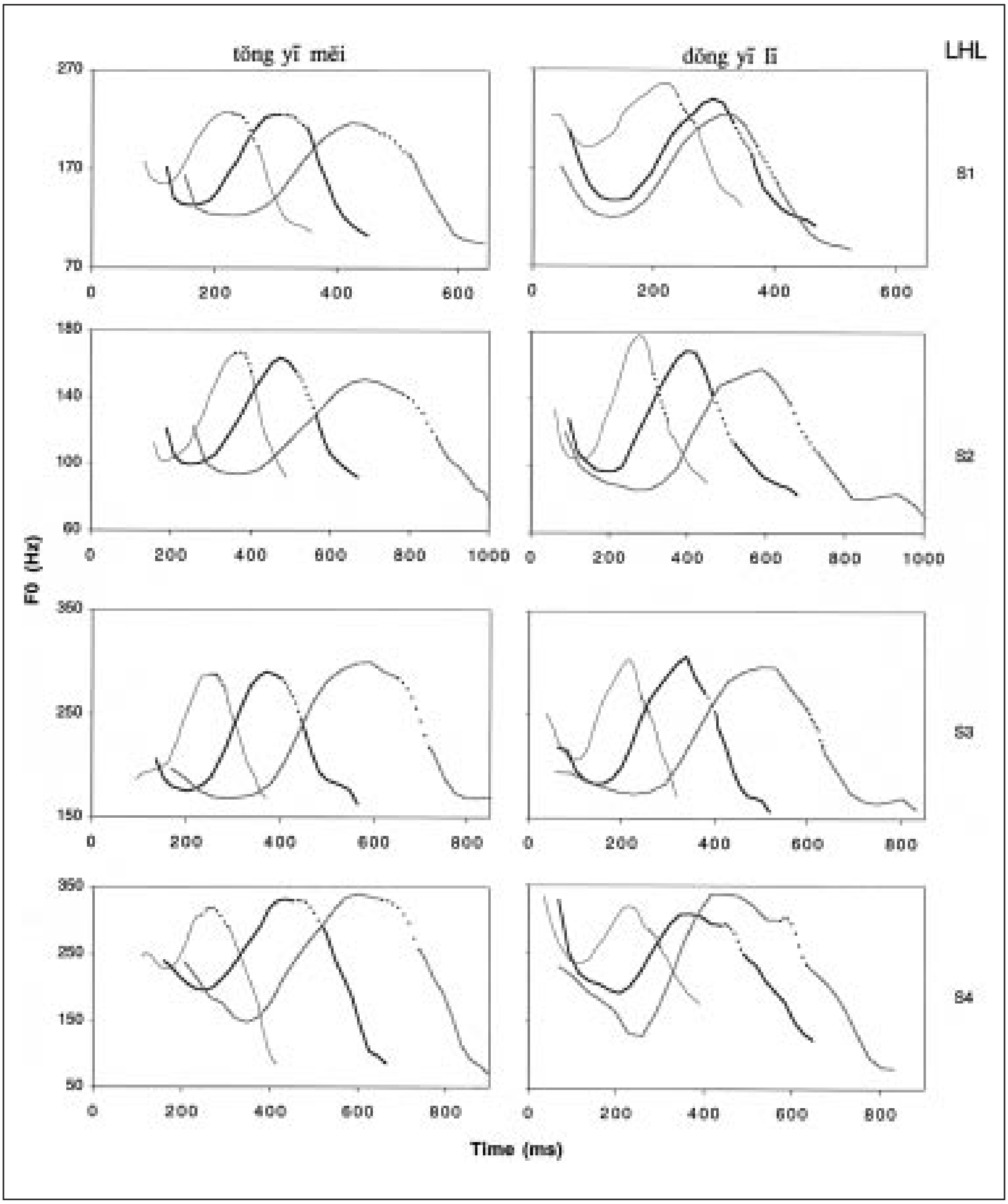
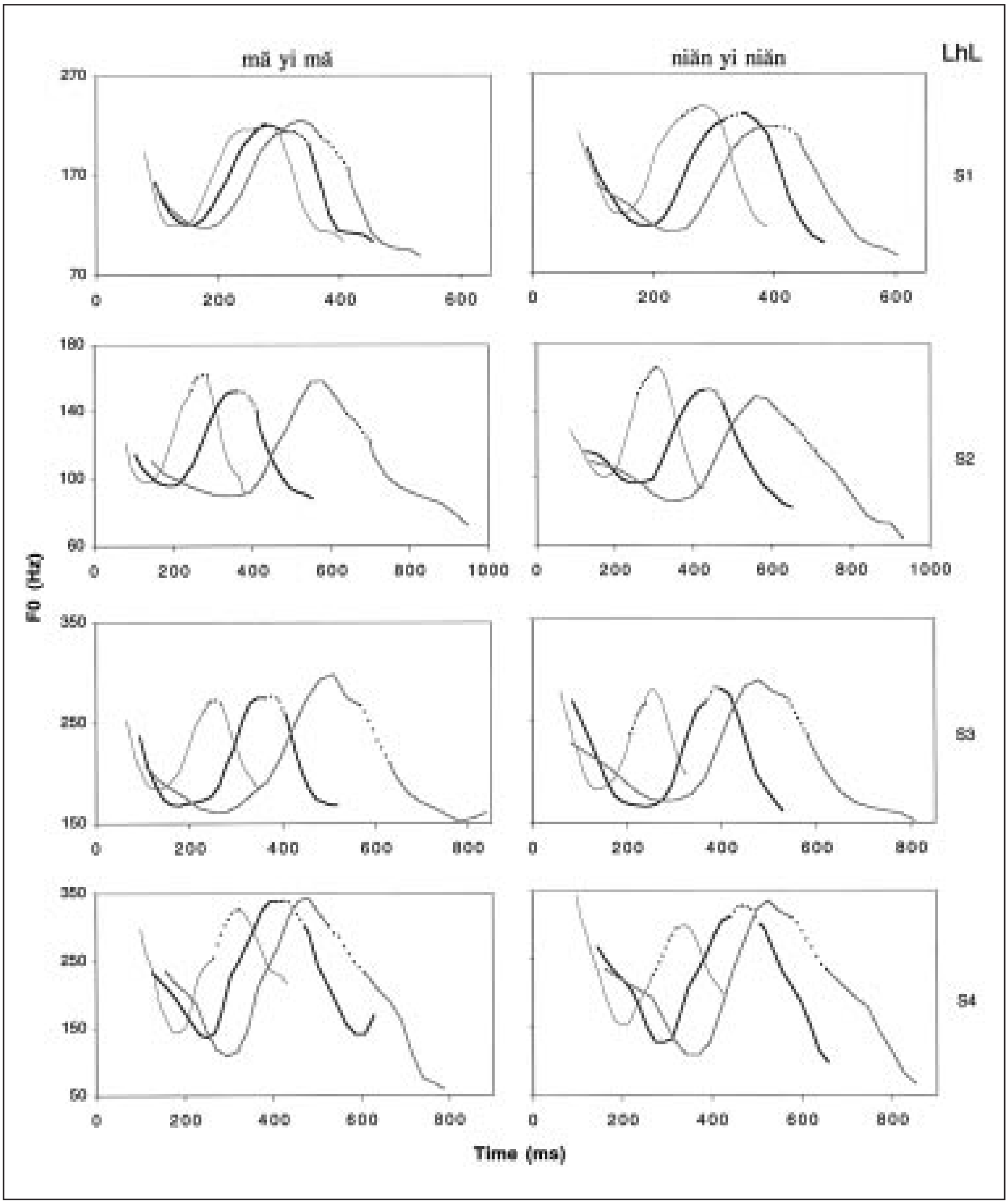


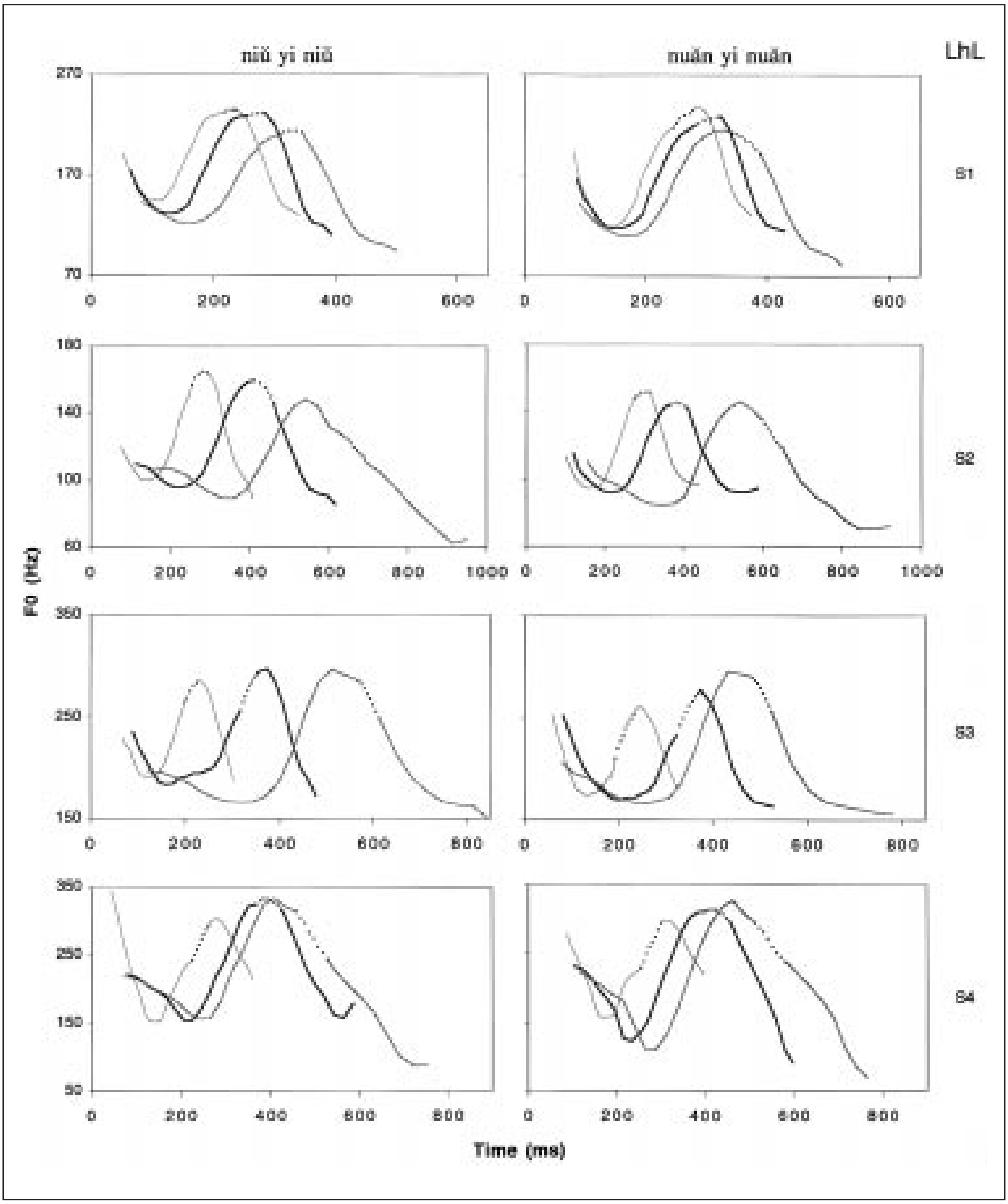
Fig. 2. Mean F_0 contours of all 4 subjects. S1 and S2 are the male subjects, and S3 and S4 are the female subjects. These F_0 contours are aligned by the onset of the first syllable in the target word/phrase, and the gap between time 0 and the beginning of each curve corresponds to the mean duration of the initial voiceless consonant of syllable 1. (Note that not all the three syllables shown on top of each column correspond to the underscored syllables in table 1. This is because, for reasons explained in the discussion about table 1 in the text, the emphasis the subjects were asked to use did not always match the target word/phrase to be analyzed.)



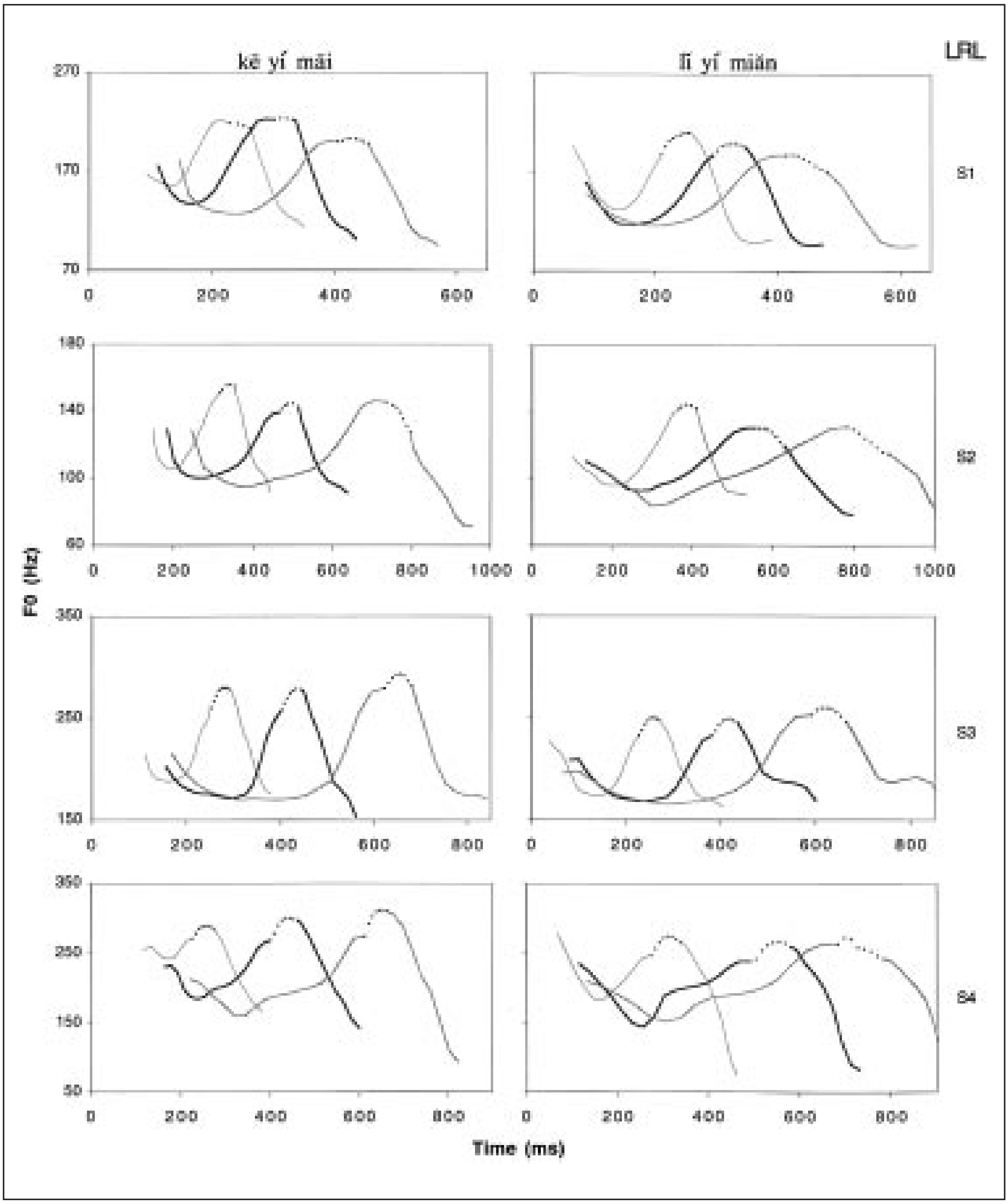
2



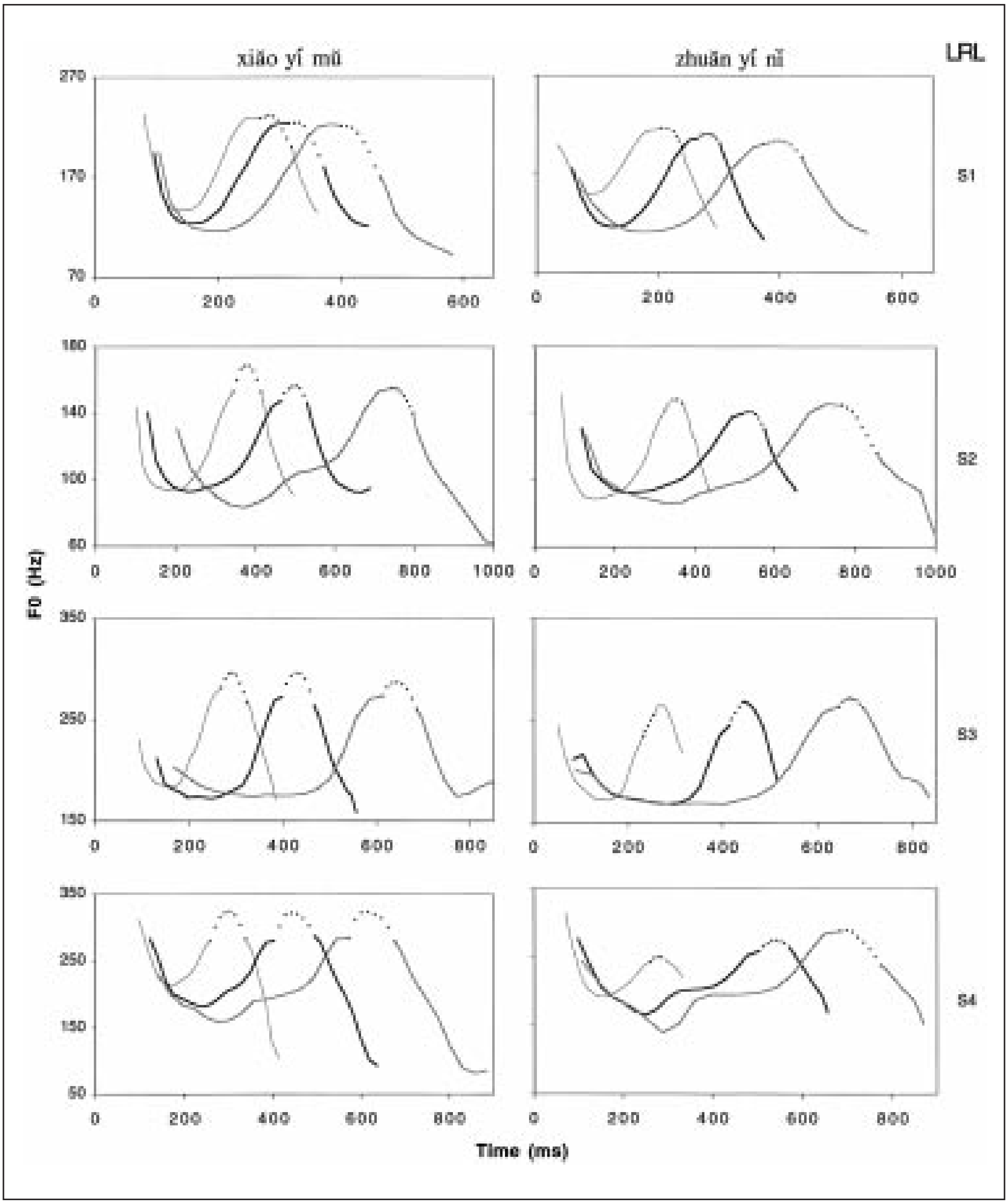
2



2



2



2

Table 2. Combined duration of syllable 1 and syllable 2 in different tone conditions and at different speech rates

Speech rate:	Fast			Normal			Slow			p		
Tone condition:	H	h	R	H	h	R	H	h	R	tone	rate	tone × rate
Subjects												
S1	221.6	236.0	220.4	297.1	281.9	282.2	418.2	351.4	394.2	0.0009	<0.0001	<0.0001
S2	324.5	248.5	325.5	472.4	381.5	482.6	723.8	627.0	749.9	<0.0001	<0.0001	0.7155
S3	242.8	200.0	233.5	395.1	333.9	386.0	630.4	533.8	612.2	<0.0001	<0.0001	0.0073
S4	238.0	241.7	241.1	441.5	392.7	431.7	612.6	510.9	629.6	<0.0001	<0.0001	<0.0001

p values are the results of two-factor ANOVAs.

fore had to drop, sometimes substantially, to reach the low pitch value required by the L tone.

The second observation is that speech rate seems to have caused clear differences in the F_0 curves. Firstly, the onset of the F_0 curves became increasingly later as speech rate slowed down, indicating that the duration of the initial consonant increased at slower speech rates. Secondly, even more duration increase can be observed in the rest of the F_0 curves. In particular, the F_0 peaks, the beginning and end of the dotted regions, and the end of the F_0 curves, all shifted rightward at slower speech rates. This indicates that the recording procedure effectively elicited systematic duration variations from the subjects. Of particular relevance to this study is whether subjects produced substantial durational differences in syllables 1 and 2. Table 2 shows the combined durations of syllables 1 and 2 in different tone and speech rate conditions for all subjects. As can be seen, the duration values are well separated across the three speech rates. The differences in duration across tone conditions are smaller. Two-factor (nonrepeated) ANOVAs were performed on duration of syllables 1 and 2 for each subject, with tone and rate as the independent variables, the combined duration of syllables 1 and 2 as the dependent variable, and repetition as the error term. The results are shown in the right section of table 2. As can be seen, the main effect of Tone is significant for all subjects. However, Student-Newman-Keuls tests found the duration difference between the H and R tones not significant for 2 subjects (S2, S4) and only marginally significant for the other 2 subjects (S1, S3). The significant interactions between tone and speech rate (except for subject S2) are mainly due to smaller duration variation in the weakened H tone than in the H and R tones at different speech rates, as can be seen in the mean duration values in table 2. This suggests that a prosodically weak syllable is probably under tighter durational constraints than the other syllables.

The third observation is that both tone and speech rate seem to have an influence on the relative location of F_0 peaks. According to the definition given in the 'Introduction', peak delay occurs whenever the F_0 peak occurs after the end of its associated syllable (i.e., syllable 2 in the target word). In figure 2, the occurrence of peak delay is indicated by the relative location of the dotted region in an F_0 curve. If it resides on the falling side of a curve, there is no peak delay, because the peak occurs before the offset of syllable 2; if it sits on top of the F_0 curve, there is moderate peak delay, because the peak occurs during the initial sonorant of syllable 3; and if it occurs on the rising side

Table 3. Mean distance between F_0 peak and offset of syllable 2 (Peak-to-C3) in different tone conditions and at different speech rates

Speech rate:	Fast			Normal			Slow			p		
Tone condition:	H	h	R	H	h	R	H	h	R	tone	rate	tone × rate
Subjects												
S1	3.5	–29.7	–15.7	7.7	–21.1	–19.7	31.9	2.8	–4.7	<0.0001	<0.0001	0.0758
S2	6.8	–37.3	–26.6	40.3	–3.4	–21.1	86.4	63.2	6.1	<0.0001	<0.0001	<0.0001
S3	6.1	–41.4	–32.2	26.0	–37.0	–35.3	70.0	35.5	–28.5	<0.0001	<0.0001	0.0003
S4	–4.6	–63.0	–36.1	19.2	–16.5	–50.9	52.8	38.2	–29.6	<0.0001	<0.0001	<0.0001

A negative value indicates that the peak occurs later than the syllable offset. p values are the results of two-factor ANOVAs.

of the curve, there is extensive peak delay, because the peak actually occurs in the vowel of syllable 3. In the LRL sequences, the F_0 peak usually occurs in syllable 3 regardless of speech rate; in the LhL sequences, the F_0 peak usually occurs in syllable 3 (the post-/yi/ syllable) at normal and fast speech rates, but in syllable 2 at slow speech rate; in the LHL sequences, the F_0 peak usually occurs in syllable 2 at normal and slow speech rates, but sometimes in syllable 3 at fast speech rate.

The fourth observation is that in ‘dǒng yīlǐ’, unlike in the other three LHL sequences, the peak in the average F_0 curves occurs before the end of syllable 2 even at fast speech rate. Relating this to the discussion in the ‘Introduction’ that /yi/ in ‘dǒng yīlǐ’ is at the left edge of a prosodic foot, it is possible that it is the difference in prosodic structure that has prevented peak delay from happening. In contrast to ‘dǒng yīlǐ’, the sequence ‘xiǎoyímǔ’, which differs from the other three LRL sequences in prosodic structure, does not show any obvious difference from them. More detailed analysis is needed to look into these two cases.

To more accurately compare the amount of peak delay in different tone and speech rate conditions, the location of F_0 peak relative to the offset of syllable 2 was measured. This location is referred to as Peak-to-C3. Table 3 shows mean values of Peak-to-C3 in different tone and speech rate conditions for all subjects. Figure 3 displays the grand averages of Peak-to-C3 across subjects. Two-factor ANOVAs were also performed on Peak-to-C3 with tone and speech rate as the independent variables. The main effects of both tone and speech rate were significant at 0.001 level for all subjects. So were all the two-way interactions, except for subjects S1 ($p = 0.0758$). Overall, as can be seen in figure 3, the peak in the mean F_0 curve is much delayed beyond the offset of syllable 2 in the R tone, but remains well within syllable 2 in the H tone across all three speech rates. The mean location of F_0 peak in weakened H, however, depends much more on speech rate: extensively delayed at normal and fast speech rates, but well within syllable 2 at slow speech rate.

Equally important as the amount of peak delay is the frequency of occurrence of peak delay. To examine that, the frequency distributions of negative Peak-to-C3 values (which indicates the occurrence of peak delay) were analyzed. Table 4 displays these distributions in different tone and speech rate conditions for all subjects. Figure 4 displays the grand average of the distributions across subjects. Two-factor ANOVAs were performed on the distribution values with tone and speech rate as the independent vari-

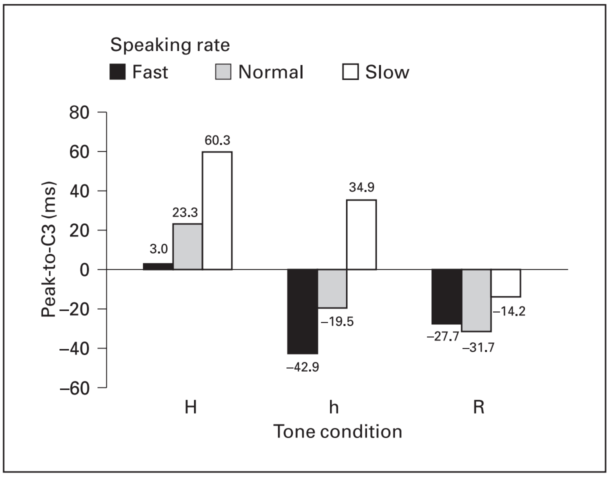


Fig. 3. Mean Peak-to-C3 values in different tone and speech rate conditions. These values are averages across all 4 subjects.

Table 4. Frequency of occurrence of peak delay (in percentage) in all tone and speech rate conditions

Speech rate:	Fast			Normal			Slow			p		
Tone condition:	H	h	R	H	h	R	H	h	R	tone	rate	tone × rate
Subjects												
S1	28.6	92.9	71.4	28.6	67.9	82.1	3.6	39.3	50	<0.0001	<0.0001	<0.1497
S2	39.3	100	96.4	0	57.1	89.3	0	0	35.7	<0.0001	<0.0001	<0.0001
S3	42.9	100	96.4	0	85.7	100	0	7.1	96.4	<0.0001	<0.0001	<0.0001
S4	71.4	100	100	28.6	78.6	100	10.7	0	92.9	<0.0001	<0.0001	<0.0001

p values are the results of two-factor ANOVAs.

ables. The results are displayed in the right section of table 4. Both table 4 and figure 4 reveal that at least for some subjects, peak delay was not an all-or-none event. It occurred more or less in all tone and speech rate conditions. At the same time, however, peak delay occurred much more freqently in R and weakened H than in the H tone, and at faster speech rates than at slower speech rates.

Peak Alignment

The analyses performed so far indicate that the amount and frequency of occurrence of peak delay varied with both tone and speech rate. In the R tone peak delay occurred frequently at all speech rates. In the H tone, although peak delay occurred in all speech rates, overall, the mean peak location remained within syllable 2, even at fast speech rate. In weakened H, peak delay occurred very frequently at normal and fast speech rates, but much less frequently at slow speech rate, and the mean peak location retreated to well within syllable 2 at slow speech rate. To further investigate the effects of tone and speech rate on peak delay, a set of regression analyses were conducted to examine the F_0 peak alignment in more detail.

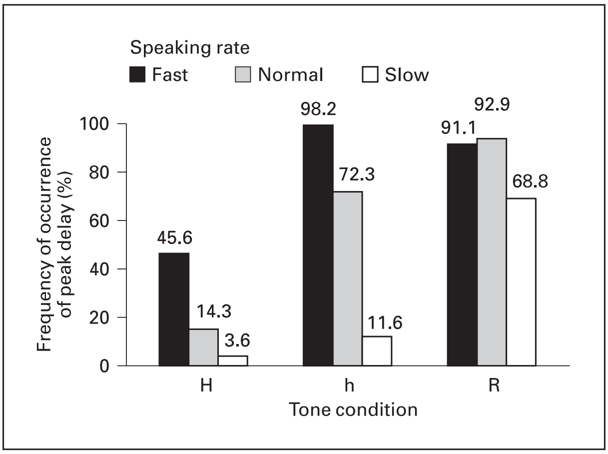


Fig. 4. Frequency distribution of peak delay in different tone and speech rate conditions. The values are averages across all 4 subjects.

First analyzed was the relationship between F_0 peak location and syllable duration. Because the boundary between syllables 1 and 2 is mostly ambiguous, as mentioned earlier, it was difficult to examine the direct relationship between peak delay and the duration of syllable 2. Assuming that the duration of both syllables 1 and 2 varied with tone and speech rate, the combined duration of the voiced portion of the two syllables (Voicing-duration 1 + 2) was used as an indicator of syllable duration. Peak-to-C3 was again used as the measurement for peak delay. In figure 5, Peak-to-C3 is plotted against Voicing-duration 1 + 2 for each subject. The linear regression equations and r^2 values for all tone conditions are displayed below each scatter plot. As can be seen in figure 5, for all subjects, the slopes of the regression lines and the r^2 values for the R tone are much smaller than those of the H and weakened H tones. This indicates that, the location of F_0 peak in the R tone did not vary much with syllable duration. In contrast, in both H and weakened H tones Peak-to-C3 varied significantly with Voicing-duration 1 + 2, indicating that syllable duration played an important role in determining their peak location. Furthermore, the slope of regression line is consistently greater in weakened H than in the H tone, indicating that there is also a certain inherent difference between the two tone conditions. Relating this difference to figure 4, it seems that it reflects the greater variation in peak location across speech rates in weakened H than in the H tone.

To test the hypothesis raised in the 'Introduction' that peak delay occurs whenever F_0 rises sharply near the end of a syllable, Peak-to-C3 was regressed over Max-velocity-to-C3, the distance between the point of maximum velocity in the F_0 curve and the offset of syllable 2 (which is also the onset of the initial consonant in syllable 3). The scatter plots and the linear regression equations for all subjects are shown in figure 6. Again, the r^2 values are high in general with the exception of those of the R tone for subjects S4 (0.164) and S2 (0.276). For S4, the low r^2 value in the R tone can be attributed to the fact that her Max-velocity-to-C3 did not vary much in this tone, as can be seen in her scatter plot in figure 6. For S2, the r^2 was probably somewhat reduced by a couple of outliers. The high overall r^2 values indicate that much of the variation in the

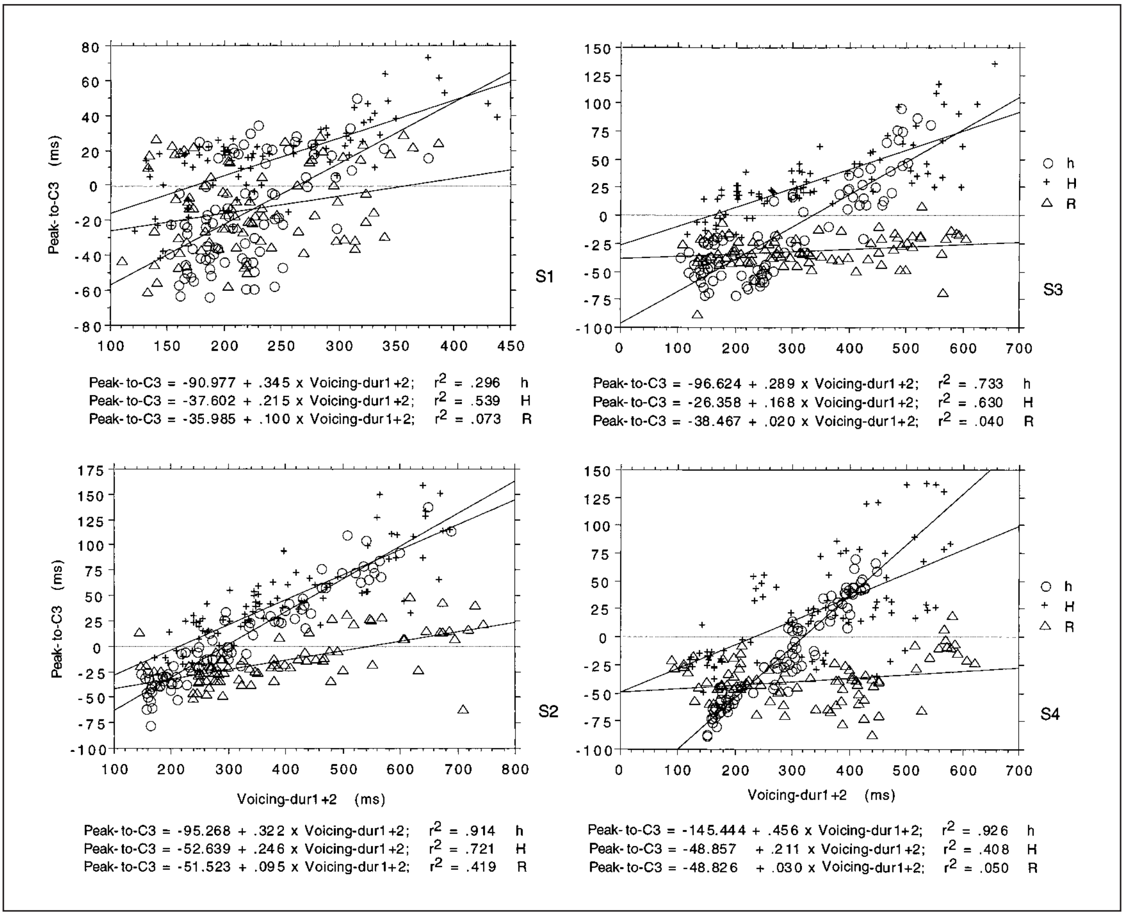


Fig. 5. Peak-to-C3 plotted against Voicing-duration 1+2 for each subject. The plotting symbols represent different tone conditions. Equations for all regression lines are displayed below each plot.

peak location relative to the offset of syllable 2 can be attributed to the location of maximum rising velocity relative to the offset of syllable 2. Once again, however, the slope of regression line is consistently higher for weakened H than for the H tone, indicating a certain difference in the inherent properties of the two tone conditions.

If the occurrence of peak delay can be at least partially predicted by the location of the sharp F_0 rise, it should be possible to assess the minimal distance between the rising contour and the syllable offset beyond which peak delay would occur. Such distance may be referred to as the critical distance. Besides the location of the maximum rising velocity shown in figure 6, the location of F_0 rise can also be 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 an F_0 curve. Linear regression equations were thus obtained using either Max-velocity-

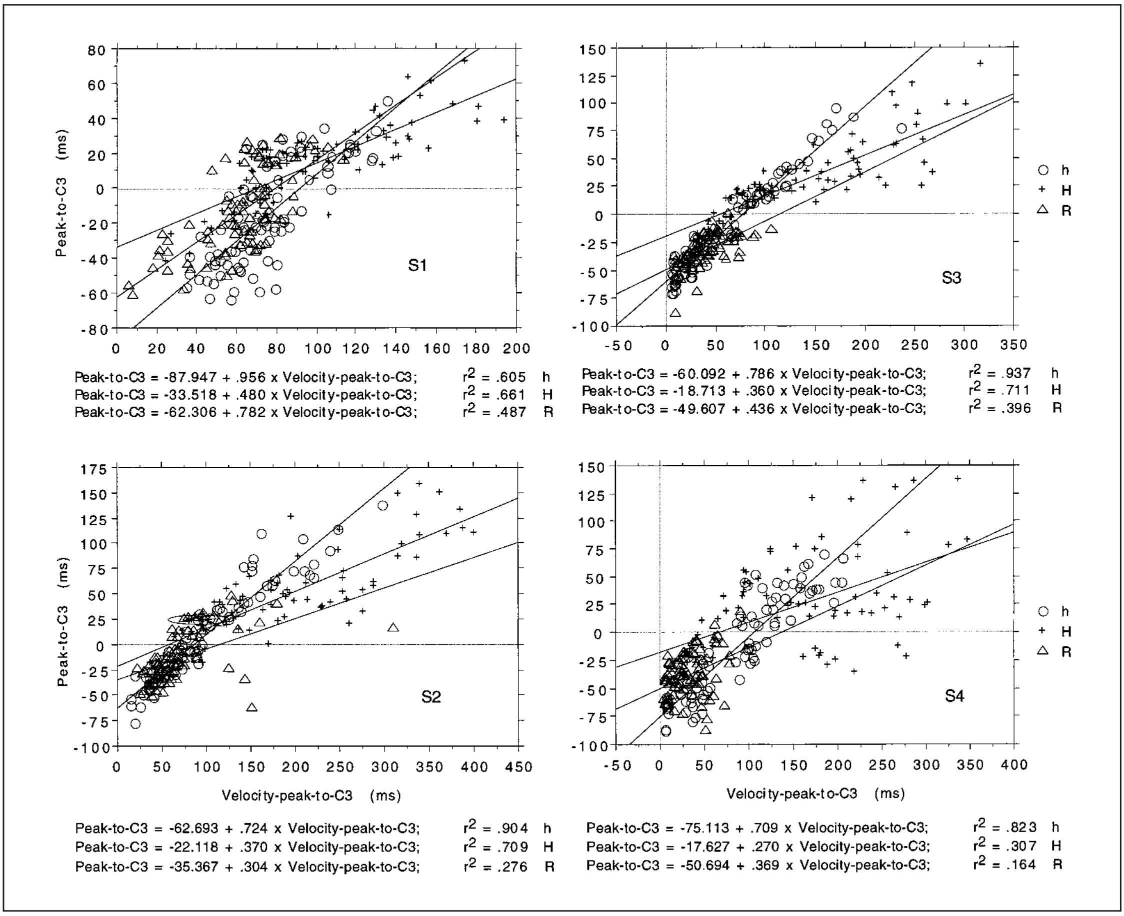


Fig. 6. Peak-to-C3 plotted against Max-velocity-to-C3 for each subject. The plotting symbols represent different tone conditions.

to-C3 or Max-acceleration-to-C3 as the predictor and Peak-to-C3 as the dependent variable. The critical distances were computed by solving the equations for the values of the predictors at Peak-to-C3 = 0. The regression equations with Max-velocity-to-C3 as predictor were already shown in figure 6. Those with Max-acceleration-to-C3 as predictor are shown in table 5. Table 6 displays the estimated critical distance values computed using the parameters shown in figure 6 and table 5.

In most cases in table 6, the critical distance for the H tone is smaller than weakened H, which is in turn smaller than the R tone. The mean critical values of Max-acceleration-to-C3 are 125.7, 144.7 and 330.2 ms for the H, weakened H and R tones, respectively. And the mean critical Max-velocity-to-C3 are 61.7, 90.3 and 111.8 ms for the H, weakened H and R tones, respectively. Some of the critical distance values for the R tone are unreasonably large (in particular, 680.6 ms for S4), given that most of the F_0 peaks were actually delayed in the R tone. It seems that the large values are due

Table 5. Parameters for linear regression equations

Tone condition	Subject	Intercept	Slope	r^2
H	S1	–40.888	0.349	0.599
	S2	–61.286	0.412	0.728
	S3	–32.505	0.33	0.657
	S4	–42.128	0.305	0.394
h	S1	–127.991	0.893	0.58
	S2	–76.788	0.544	0.73
	S3	–89.132	0.672	0.75
	S4	–126.869	0.784	0.898
R	S1	–53.079	0.311	0.274
	S2	–26.813	0.087	0.05
	S3	–66.878	0.415	0.438
	S4	–45.599	0.067	0.038

They were obtained with Max-acceleration-to-C3 as the predictor and Peak-to-C3 as the dependent variable. r^2 values for all regression analyses are shown in the right most column.

Table 6. Critical values of Max-acceleration-to-C3 and Max-velocity-to-C3 for each tone produced by each subject

Estimated critical value:	Max-acceleration-to-C3			Max-velocity-to-C3		
Tone condition:	H	h	R	H	h	R
Subjects						
S1	117.2	143.3	170.7	69.8	92.0	79.7
S2	148.8	141.2	308.2	59.8	86.6	116.3
S3	98.5	132.6	161.2	52.0	76.5	113.8
S4	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

The critical values were computed using the linear regression equations in the form of Peak-to-C3 = a + b × Predictor, where Predictor is either Max-acceleration-to-C3 (table 5) or Max-velocity-to-C3 (fig. 6). Each critical value was computed by solving the respective equation at Peak-to-C3 = 0.

to the shallow slope of the respective regression lines. Apparently, these values should not be considered as the true critical distances for the R tone, because peak alignment in R did not vary much with syllable duration in the first place, as shown in figure 5 and discussed earlier. In fact, for 3 of subjects, Max-acceleration-to-C3 and Max-velocity-to-C3 remain close to the critical values estimated from the H and weakened H tones. Table 7 displays the means of these two values for all subjects across tone and speech rate conditions. As can be seen in table 7, the largest mean Max-acceleration-to-C3 is 128.3 ms for the R tone, while the smallest mean critical Max-acceleration-to-C3 is 125.7 ms in table 6. The largest mean Max-velocity-to-C3 in the R tone is 70.5 ms in table 7, while the smallest mean critical Max-velocity-to-C3 is 61.7 ms in table 6. It

Table 7. Mean Max-acceleration-to-C3 and mean Max-velocity-to-C3 values

Speech rate:	Fast			Normal			Slow		
Tone condition:	H	h	R	H	h	R	H	h	R
Max-acceleration-to-C3									
S1	115.3	108.6	102.2	152.1	122.5	130.1	208.7	144.8	150.1
S2	166.0	91.0	131.7	249.6	142.1	168.7	355.7	231.7	146.4
S3	112.0	71.5	73.6	187.8	94.8	75.6	283.1	167.5	103.3
S4	106.1	81.2	90.1	218.0	149.2	99.3	311.6	202.4	113.4
Mean	124.9	88.1	99.4	201.9	127.2	118.4	289.8	186.6	128.3
Max-velocity-to-C3									
S1	70.1	60.9	53.1	89.9	70.6	58.5	139.0	94.1	76.2
S2	84.7	39.1	46.0	171.4	83.0	66.8	284.0	168.6	99.8
S3	67.6	18.7	27.9	121.5	35.2	33.4	230.4	120.9	59.9
S4	52.0	27.2	22.6	151.8	86.8	27.4	242.5	145.3	46.1
Mean	68.6	36.5	37.4	133.7	68.9	46.5	224.0	132.2	70.5

therefore appears that, given the smallest critical distances in table 6, it is natural that peak delay should frequently occur in the R tone: because its final F_0 rise consistently occurs near or within the critical distance from the syllable offset.[5]

Finally, to examine whether differences in prosodic structure had any effect on peak delay, individual sequences in the LHL condition were compared for their values of Peak-to-C3 and frequency of occurrence of peak delay. Only the latter revealed any consistent difference, however, as shown in figure 7. As can be seen in figure 7, while the frequency of occurrence of peak delay of most sequences varied much across subjects, 'dong yili' is the only one for which peak delay is consistently missing at all speech rates across all subjects. Because the experiment was not designed for testing the relation between peak delay and microscopic prosodic structures, however, the difference observed in figure 7 could not be substantiated statistically. Similar comparisons were also made among all the LRL sequences, but no consistent differences were observed.

Discussion

The results of the various analyses on the F_0 contours of the utterances recorded in the present study have revealed interesting information about peak delay in Mandarin. The magnitude as well as frequency of occurrence of peak delay were found to vary across both tone and speech 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

[5] Note that this is not an explanation for why the rise occurs consistently late in a syllable for the R tone. The late rise in R, as explained earlier, is necessitated by its underlying rising contour. The comparison of the values in table 7 with those in table 6 for the R tone is only a confirmation that if the rise consistently occurs near the syllable offset, peak delay is likely to occur frequently.

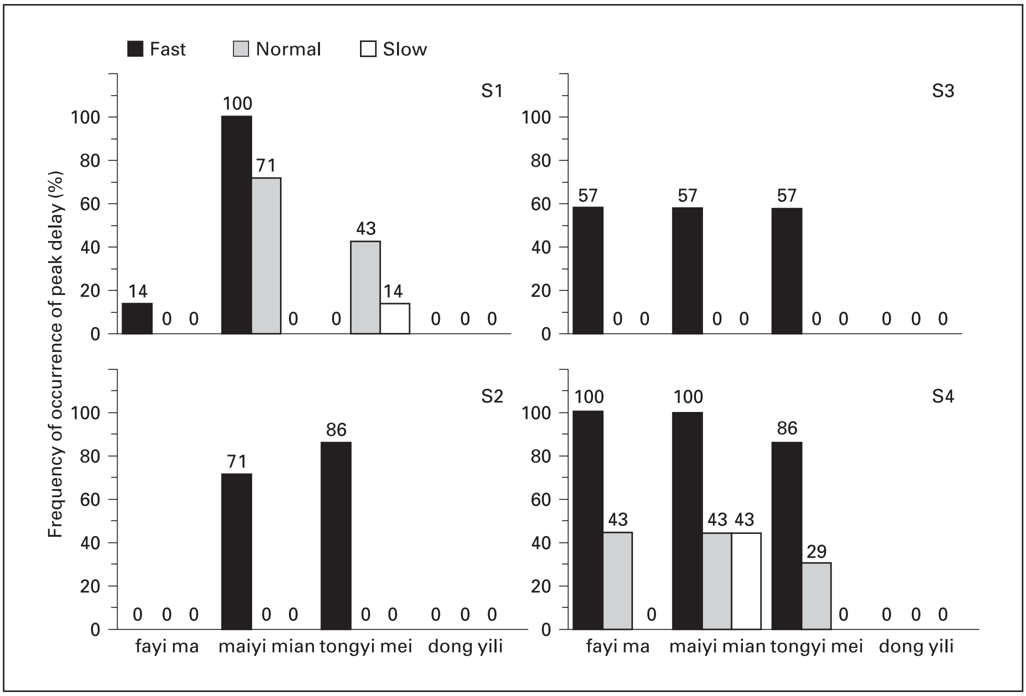


Fig. 7. Frequency of occurrence of peak delay in individual LHL sequences. Note the consistent absence of peak delay in the sequence 'dong yili' at all three speech rates across subjects.

more likely to occur if the F_0 rise was close to the offset of the tone-carrying syllable. An attempt was also made to assess the critical distance between the F_0 rise and the syllable offset at which peak delay may occur. According to these assessments, peak delay is likely to occur if the F_0 rise starts at about 125 ms or less before the end of a syllable, even if the tone is H, in which peak delay is in general the least likely to occur.

General Discussion

The findings of the present study seem to support the 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 as proposed in Xu and Wang [in press]. To illustrate the hypothesis, figure 8 shows a schematic depiction of peak alignments in the F, H and R tones in Mandarin at normal speech rate and those of the H and R tones at fast speech rate. As shown in the upper panel of figure 8, 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 the F tone the peak is by necessity relatively early due

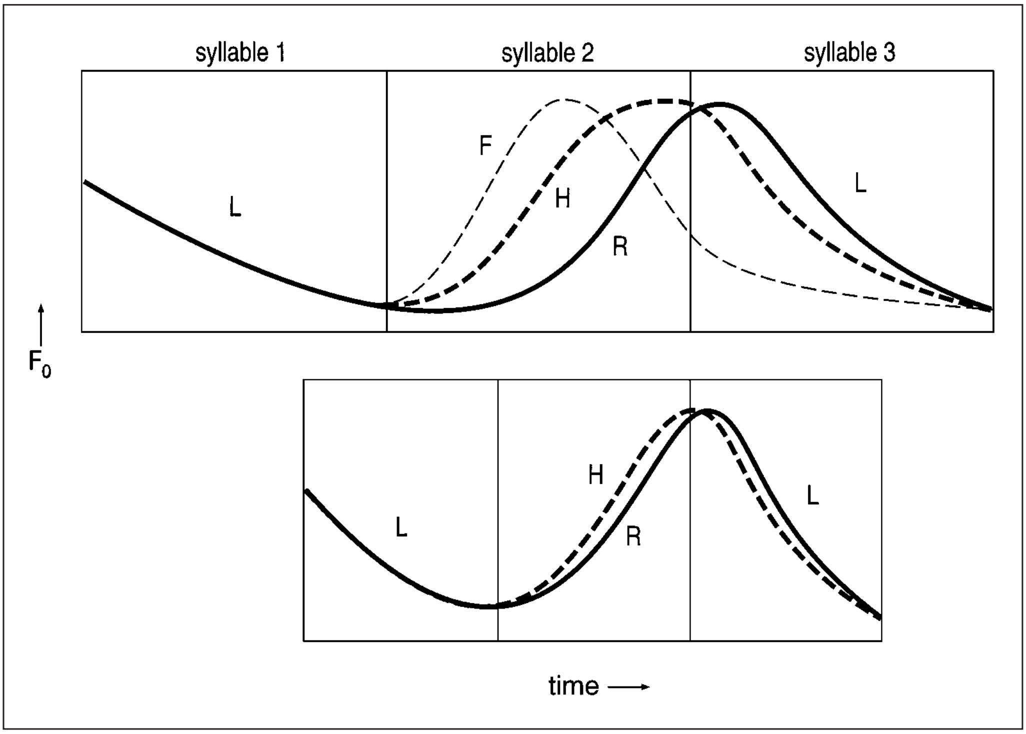


Fig. 8. An illustration of peak alignment in the F, H and R tones in Mandarin at normal speech rate (upper panel), and peak alignment in the H and R tones at fast speech rate (lower panel). See text for discussion.

to its underlying target [fall], whereas the peak in the R tone is relatively late due to its underlying target [rise]. For the H tone, since its target is presumably a static [high] [Xu and Wang, in press], the F_0 contour should tend to level off whenever possible, and its F_0 peak would occur between those of the F and R tones.

Figure 8 also illustrates what happens to F_0 peaks in the R and H tones when syllable duration varies at different speech rates. In the H tone, at normal (or slow) speech rate, there is plenty of time for the F_0 contour to level off before the syllable offset, hence peak delay does not usually occur. In contrast, in the R tone a sharp rise is required by the underlying pitch target, and this rise is always implemented right before the syllable offset [Xu, 1998]. Since it takes time for the larynx to terminate the rise right at the syllable offset, the actual peak usually occurs soon after the end of the syllable. At fast speech rate, the rising slope in the H tone is pushed up to the end of the syllable. As a result, the peak occurs either very close to the syllable offset, or occasionally after the syllable boundary. Meanwhile, since the sharp rise in the R tone is usually already implemented in the last portion of a syllable at normal speech rate, less change in peak delay happens in the R tone at fast speech rate.

In short, a sharp rise in F_0 may occur close to the syllable offset for different reasons. It may occur there due to the requirement of the underlying [rise], which is always implemented most fully by the end of the syllable regardless of syllable duration. Or it may occur there when the H tone is preceded by the L tone and the syllable duration is short. But whenever the rise occurs close to the syllable offset, peak delay is likely to occur.

One could argue that it would be theoretically just as adequate to treat F_0 peaks and their alignment as the primitive targets, and view F_0 contours as the natural consequences of peak alignment. Indeed Gårding [1977] shows 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. Ladd [1983, p. 732] demonstrates that a feature [delayed peak] can be used to distinguish two kinds of falls in English both as HL but with different peak locations: 'plain fall is [– delayed peak], while scooped fall is [+ delayed peak]'. Gårding and Zhang [1986] and Gårding [1987] also argue that turning point can be used to distinguish Mandarin tones. Arvaniti et al. [1998] while reporting a peak alignment pattern in Greek somewhat similar to that of the R tone in Mandarin, i.e., having an F_0 peak nearly always occurring right after the stressed syllable,[6] argue that such alignment 'is difficult to accommodate in a theory that views pitch movement as the primes of intonational structure'. Indeed, if observed F_0 peaks and their alignment are viewed as equivalent to the underlying tonal or intonational targets, in certain cases, such as the antepenultimate lexical stress in Greek, prenuclear H* in English, and the R tone in Mandarin (when applicable), F_0 peaks do seem to display a very consistent characteristic, i.e., occurring either right after the vowel onset in the postaccentual syllable [Greek, Arvaniti et al., 1998], in the later portion of the consonant immediately after the stressed syllable [English, Ladd et al., 1999], or right after the R-carrying syllable [Mandarin, Xu, 1998, 1999, present study].

However, as found by Xu [1997, 1999], at least in Mandarin, the R tone does not even have to have a peak. When it is followed by the H tone there is often no peak around the syllable boundary. When the R tone is followed by the F tone, the peak usually occurs rather late in the F-carrying syllable, but that peak is more likely to be associated with the F rather than the R tone, as observed in Xu [1999]. Similarly, there may not be a peak in the F tone when it is preceded by the H tone, and there may not be a peak in the H tone if it happens to be in the tonal context of H_H or R_H [Xu, 1999]. The only thing really consistent about a tone across various tonal context, speech rate, and focus conditions in Mandarin is the F_0 contour in the later portion of its host syllable. As argued in Xu [1999] and Xu and Wang [in press], the acoustic manifestation of a tone does not always directly reflect its underlying pitch target. Rather, it reflects the speaker's implementation of the underlying pitch target within the time frame of its carrying segmental unit (the syllable, in the case of Mandarin). In the case of the R tone, the underlying pitch target is [rise] and hence this target is implemented the fullest by the end of the R-carrying syllable. When the R tone is followed by the L tone, the implementation of the L tone (whose pitch target is [low]) starts as soon as the R-carrying syllable ends. However, because it takes time for the larynx to terminate a sharp rise, the actual F_0 turning point occurs somewhat *after* the R-L syllable boundary.

[6] As pointed out by Ladd in his review of our manuscript, the detailed peak alignment in Greek reported by Arvaniti et al. [1998] is not exactly the same as that in Mandarin R tone. In Greek prenuclear accent the peak is on average 15–20 ms into the following unstressed vowel, whereas the peak of the rise in Mandarin mostly occurs just inside the following consonant.

Thus in this particular tone sequence (R-L), the highly consistent alignment of F_0 peak is due to the consistent alignment of the two tones with their respective host syllables. What the present study has found further is that if the rise in the H tone due to the preceding L tone is pushed up to the offset of the H-carrying syllable at fast speech rate, the turning point in F_0, which at normal rate usually occurs before the syllable offset, can also occur after the end of the syllable. This finding provides additional support for the hypothesis that articulatory constraints interact with underlying pitch targets in determining the shape and alignment of *the surface* F_0 contours.

For Mandarin, therefore, the occurrence and alignment of F_0 peaks, as well as the occurrence of peak delay, may be accounted for by the interactive contribution of underlying pitch targets and their articulatory implementation. In contrast, treating the F_0 peaks as the primitive targets and using them to predict the shape of F_0 contours in Mandarin would seem both effortful and circuitous. For other languages, there might be a more even balance between the two approaches at the descriptive level, as has been demonstrated by Gårding [1977], Ladd [1983], Silverman and Pierrehumbert [1990] and many other studies. Still, findings about Mandarin F_0 contour alignment may offer a new perspective for understanding F_0 contour alignment in other languages. For example, to determine for a particular language if a tone, whether lexical or accentual, consists of a single dynamic target or a sequence of static targets, it may help to examine not only critical F_0 points such as peaks and valleys, but also the exact shape of the F_0 contours as indicated by the velocity and acceleration of the F_0 change, and possibly other measurements as well. Furthermore, it may also be helpful to make explicit assumptions about the tonal context for the tone under scrutiny, knowing that it always takes time for the speaker to complete a pitch change [Ohala and Ewan, 1973; Sundberg, 1979]. Finally, given that the F_0 contour of a tone varies extensively in different tonal contexts, observed F_0 contours probably should be treated not as the equivalent of the underlying tonal targets themselves, but rather as the realizations of the underlying targets in specific tonal contexts under specific articulatory constraints.

The present study, however, has not exhausted all the possible factors that may contribute to the occurrence of peak delay in Mandarin. For example, the difference within the LHL sequence, as shown in figure 7, indicates that prosodic structures in Mandarin may also influence the location of F_0 peaks. Such an influence may be exerted in two ways – by affecting syllable duration and/or by varying the amount of effort used in approaching a pitch target. As indicated by the general finding of the present study, a longer syllable duration may better allow a peak to be reached within the syllable boundary. Similarly, greater efforts may also increase the likelihood that the peak is realized before the syllable offset. It awaits future studies, however, to further explore the relation between peak delay and prosodic structure and possibly other factors as well.

Conclusions

Observed variability in the surface alignment of F_0 contours and segmental units may make it appears as if phonological pitch units such as tone and pitch accent are in general not strictly aligned to their lexically or prosodically associated segmental units such as the syllable. Recent studies of Mandarin tones have found that much of the variability in the shape and alignment of F_0 contours in Mandarin can be attributed to the interaction among the underlying pitch targets, tonal contexts and articulatory constraints [Xu, 1997, 1998, 1999]. The present study further investigated such interaction by examining the phenomenon of peak delay. Specifically, the prediction was tested that if a sharp pitch rise occurs near the syllable offset, regardless of the cause of the rise, the F_0 peak associated with that syllable is likely to occur in the following L-tone-carrying syllable. In particular, if the duration of a LHL tone sequence is sufficiently shortened, a sharp rise in F_0 may occur near the end of the H-carrying syllable, and the peak associated with the H tone may be delayed into the second L-carrying syllable, much like the peak delay in a LRL sequence which occurs regularly at normal speech rate. The results of the study in general verified this prediction. Peak delay occurred not only regularly in the LRL sequences at all speech rates, but also quite often in the LHL sequences at fast speech rate. This finding therefore provides support for the interactive account for the variability of F_0-contour alignment in Mandarin. Furthermore, the finding provides support for the proposal [Xu and Wang, in press] that there may be no actual underlying misalignment between pitch units and segmental units in Mandarin, and possibly in many other languages either, and that the observed surface misalignment may be a consequence of implementing pitch targets and their underlyingly aligned segmental units in certain prosodic contexts under certain articulatory constraints.

The results of the study also indicate that the prosodic structure of an utterance may also play a role in shaping and aligning F_0 contours. The precise nature of the role, however, awaits further investigation in future studies. Finally, although an attempt was made to assess the 'critical distance' for the occurrence of peak delay, how it relates to actual physiological limitation of the larynx is not yet quite clear. Further studies are also needed to look into it.

Acknowledgments

I would like to thank Bob Ladd and an anonymous reviewer for their insightful comments on an earlier version of this paper. My thanks also go to Xuejing Sun who helped with data collection and analysis for the study. Part of the results of this study was presented at the 14th International Congress of Phonetic Sciences, San Francisco. This study was supported in part by NIH grant DC03902.

References

Arvaniti, A.; Ladd, D.R.; Mennen, I.: Stability of tonal alignment: the case of Greek prenuclear accents. J. Phonet. *36:* 3–25 (1998).

Chao, Y.R.: A Grammar of Spoken Chinese (University of California Press, Berkeley, CA 1968).

Gårding, E.: The importance of turning points; in Hyman, Studies in stress and accent, pp. 27–35 (Department of Linguistics, University of Southern California, Los Angeles 1977).

Gårding, E.: Speech act and tonal pattern in Standard Chinese. Phonetica *44:* 13–29 (1987).

Gårding, E.; Zhang, J.: Tone 4 and Tone 3 discrimination in Modern Standard Chinese. Lang. Speech *29:* 281–293 (1986).

Grimm, C.: Pitch accent in Oneida. Proc. Annu. Meet. Ling. Soc. Am., Chicago 1997.

Hombert, J.-M.: Consonant types, vowel quality, and tone; in Fromkin, Tone: linguistic survey, pp. 77–111 (Academic Press, New York 1978).

Jin, S.: An Acoustic Study of Sentence Stress in Mandarin Chinese. PhD diss. Ohio State University (1996).

Jong, K. de: Initial tones and prominence in Seoul Korean. Ohio State Univ. Working Papers Ling. *43:* 1–14 (1994).

Kim, S.-A.: Positional effect on tonal alternation in Chichewa: phonological rule vs. phonetic timing. Proc. Annu. Meet. Chicago Ling. Soc., vol. 34, pp. 245–257, Chicago 1999.

Ladd, D.R.: Phonological features of intonational peaks. Language *59:* 721–759 (1983).

Ladd, D.R.; Faulkner, D.; Faulkner, H.; Schepman, A.: Constant 'segmental anchoring' of F_0 movements under changes in speech rate. J. acoust. Soc. Am. *106:* 1543–1554 (1999).
Lehiste, I.; Peterson, G.E.: Some basic considerations in the analysis of intonation. J. acoust. Soc. Am. *33:* 419–425 (1961).
Ohala, J.J.; Ewan, W.G.: Speed of pitch change. J. acoust. Soc. Am. *53:* 345A (1973).
Prieto, P.; Santen, J. v.; Hirschberg, J.: Tonal alignment patterns in Spanish. J. Phonet. *23:* 429–451 (1995).
Santen, J.P.H. v.; Hirschberg, J.: Segmental effects on timing and height of pitch contours. Proc. Int. Conf. on Spoken Lang. Processing, vol. 94, pp. 719–722 (1994).
Shih, C.: The prosodic domain of tone Sandhi in Chinese; PhD diss. University of California, San Diego (1986).
Shih, C.; Sproat, R.: Variations of the Mandarin rising tone. Proc. IRCS Workshop on Prosody in Natural Speech No. 92-37, pp. 193–200 (Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia 1992).
Silverman, K.E.A.; Pierrehumbert, J.B.: The timing of prenuclear high accents in English; in Kingston, Beckman, Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech, pp. 72–106 (Cambridge University Press, Cambridge 1990).
Speer, S.R.; Shih, C.; Slowiaczek, M.L.: Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. Lang. Speech *32:* 337–354 (1989).
Steele, S.A.: Nuclear accent F_0 peak location: effects of rate, vowel, and number of following syllables. J. acoust. Soc. Am. *80:* S51 (1986).
Sundberg, J.: Maximum speed of pitch changes in singers and untrained subjects. J. Phonet. *7:* 71–79 (1979).
Xu, Y.: Pǔtōnghuà yīnlián de shēngxué Yǔyīnxué tèxìng (The acoustic-phonetic characteristics of junctures in Mandarin). Zhōngguó Yǔwén 353–360 (1986).
Xu, Y.: Contextual tonal variations in Mandarin. J. Phonet. *25:* 61–83 (1997).
Xu, Y.: Consistency of tone-syllable alignment across different syllable structures and speaking rates. Phonetica *55:* 179–203 (1998).
Xu, Y.: Effects of tone and focus on the formation and alignment of F_0 contours. J. Phonet. *27:* 55–105 (1999).
Xu, Y.; Wang, Q.E.: Pitch targets and their realization: evidence from Mandarin Chinese. Speech Communication (in press).