Journal of Phonetics (1999) **27**, 55–105
Article No. jpho.1999.0086
Available online at http://www.idealibrary.com on IDEAL®

Effects of tone and focus on the formation and alignment of f_0 contours

Yi Xu

Department of Communication Sciences and Disorders, Speech and Language Pathology, Northwestern University, 2299 North Campus Drive, Evanston, IL 60208, U.S.A.

Received 10th February 1998, revised and accepted 14th January 1999

The present study examines how lexical tone and focus contribute to the formation and alignment of f_0 contours in speech. This was done through an investigation of f_0 contour formation in short Mandarin sentences. These sentences all consisted of five syllables with varying tones on the middle three syllables. The sentences were produced by eight Mandarin speakers with four different focus patterns: focus on the first, second, or last word, or with no narrow focus. The f_0 patterns of these sentences were examined through point-by-point f_0 tracing, graphical comparison of averaged f_0 contours, f_0-contour-syllable alignment analysis, and analysis of maximum, minimum f_0, and slope of f_0 contours. The results indicate that (a) while the lexical tone of a syllable is the most important determining factor for the local f_0 contour of the syllable, focus extensively modulates the global shape of the f_0 curve, which in turn affects the height and even the shape of local contours; (b) the tones of adjacent syllables also extensively influence both the shape and height of the f_0 contour of a syllable, with the preceding tone exerting more influence than the following tone; (c) despite extensive variations in shape and height, the f_0 contour of a tone remains closely aligned with the associated syllable; and (d) both focus and tonal interaction may generate substantial f_0 decline over the course of an utterance. These findings seem to be able to reduce the unpredictability in the formation and alignment of f_0 contours in speech.

1. Introduction

The f_0 (fundamental frequency) curve of a speech utterance is known to be a major acoustic manifestation of suprasegmental structures such as tone, pitch accent, and intonation. However, just as vowels and consonants do not have invariant spectrographic representations, suprasegmental structures may not have a one-to-one correspondence with observed f_0 patterns. Surface f_0 contours do not necessarily resemble the underlying suprasegmental structures, because many variations are introduced during the implementation of these structures. It is therefore often difficult to understand f_0 patterns through direct observation. For example, it has been observed that there is

0095-4470/99 010055 + 51 $30.00/0

a tendency for f_0 to gradually decline over the course of an utterance. The phenomenon is known as declination (Cohen & 't Hart, 1965; Cohen, Collier & 't Hart, 1982) and has been reported in many languages (Pike, 1945; Maeda, 1976; Cooper & Sorensen, 1981; Ohala, 1990; Shih, 1997). After decades of research, however, it is still unresolved as to whether declination is a functionally distinct intonation pattern ('t Hart & Collier, 1975; Cooper & Sorensen, 1977, 1981; Ohala, 1978, 1990) or a byproduct of utterance production (Collier, 1975, 1984, 1987; Lieberman, 1967; Maeda, 1976; Titze & Durham, 1987). This is probably because, as pointed out by Liberman and Pierrehumbert (1984), most of the research on phenomena like declination has been based on observations of global f_0 patterns without adequate analysis of the local prosodic structures. For intonation, however, the difficulty often is that the underlying specifications of the local components are hard to estimate. For one thing, they do not usually occur in isolation; for another, they may not be independent of the global patterns under scrutiny.

To better understand the surface f_0 patterns, it is therefore desirable to find prosodic structures that are relatively independent of intonation and whose underlying specifications are relatively well understood. These structures can then be examined for their realization as f_0 contours under the influence of various factors. Lexical tones in tone languages may serve this purpose. The underlying specifications of lexical tones can be estimated relatively independent of intonation by keeping intonation constant. This was done in many early tone studies, mostly on Asian tone languages (Bai, 1934; Chao, 1948, 1956, 1968; Abramson, 1962, 1976, 1978; Lin, 1965, 1988; Howie, 1976; Chuang, Hiki, Sone & Nimura, 1971; Ho, 1976). As found by these studies, lexical tones are specified mainly in terms of the height and shape of pitch contours. For example, it has been well established that the four lexical tones (not including the neutral tone) in Mandarin — H (also known as Tone 1), R (Tone 2), L (Tone 3), and F (Tone 4) — have the pitch contours high-level, mid-rising, low-dipping, and high-falling, respectively (Chao, 1948, 1956, 1968; Lin, 1965, 1988; Howie, 1976; Chuang *et al.*, 1971; Ho, 1976). These tones are therefore known as contour tones. In contrast to contour tones, tones in many non-Asian tone languages are known as register tones (Pike, 1948), because they each have a single underlying pitch specification, such as H (high), M (mid), or L (low). Given these underlying pitch specifications, it is possible to examine how lexical tones interact with various factors to form surface f_0 contours.

One of the important factors is tonal context. In African tone language research it has been known for a long time that the f_0 height and contour of a tone are affected by adjacent tones (Hyman, 1973; Hyman & Schuh, 1974). In a HLH sequence, for example, the f_0 height of the second H is lower than that of the first H, presumably because it is lowered by the preceding L. This phenomenon is known as downstep, and it has been found in many African tone languages (Stewart, 1965, 1983; Meeussen, 1970; Hyman, 1973). Interestingly, downstep has also been reported in a number of non-tone languages (e.g., Pierrehumbert, 1980, for English; Poser, 1984, and Pierrehumbert & Beckman, 1988, for Japanese; and Prieto, Shih & Nibert, 1996, for Spanish) and in a contour tone language (Mandarin, as reported by Shih, 1988). When downstep occurs repeatedly, it may generate an overall downward f_0 tilt over the entire course of an utterance. Since declination is also known as an overall f_0 downtrend in an utterance, there might be certain similarities or possibly some overlap between the two phenomena. Indeed, Pierrehumbert (1980) and Liberman and Pierrehumbert (1984) reported that much of the time-dependent lowering in English could be accounted for by a downstep model with an

accent-by-accent decay of a constant ratio. Prieto *et al.* (1996) also reported that in Mexican Spanish the heights of successive f_0 peaks could be predicted by exclusively applying a local downstep ratio. Based on this finding, they suggested that declination is probably equivalent to a series of downsteps.

A deeper understanding of downstep therefore seems important for understanding downtrend in general. A study by Shih (1988) on contextual tonal variations in Mandarin reported some findings that seem critical for understanding downstep. Shih found that the amount of f_0 lowering in a tone due to downstep differed depending on which of the four tones preceded it. The f_0 was found to be lowered the most when the tone was preceded by the L tone, but only moderately lowered when preceded by the R and F tone. There was no lowering by the preceding H tone. Although difficult to accommodate using a constant downstep ratio, such "different downstep" (Beckman, 1995, p. 106) may be viewed as directly related to the f_0 height of the tone that triggers the lowering, as suggested by Shih (1988). When produced in context, Mandarin L tone has the lowest minimum f_0, R and F tones have moderate minimum f_0, and H tone has the highest minimum f_0. Such an account is supported by other studies that have investigated contextual variations of contour tones in Mandarin, Thai, and Vietnamese (Han & Kim, 1974; Xu, 1993, 1994, 1997; Gandour, Potisuk & Dechongkit, 1994).

Furthermore, it has been found that contextual tonal variations consist of not only the carryover assimilatory effect discussed above, but also an anticipatory dissimilatory effect (Gandour, Potisuk, Dechongkit & Ponglorpisit, 1992; Gandour *et al.*, 1994; Xu 1993, 1997). The anticipatory effect, also known as anticipatory raising or regressive H-raising, refers to the phenomenon that the f_0 height of a tone is raised when followed by a L tone. Besides Thai and Mandarin, anticipatory raising has also been reported for a number of African tone languages (Laniran, 1992, for Yoruba; Hyman, 1993, for Enginni, Mankon, and Kirim; and Laniran & Gerfen, 1997, for Igbo). In addition, Laniran (1992) and Laniran and Clements (1995) further suggest that anticipatory raising may in fact be the real mechanism underlying downstep.

Taken together, the findings in tone language research suggest that there is much to be learned about contextual tonal variations and their contribution to the local as well as global f_0 patterns of an utterance. In particular, the scope and nature of the anticipatory and carryover effects need to be better understood. Xu (1993, 1997) and Gandour *et al.* (1994) investigated these effects, but the scope of the effects examined was limited. The kind of downstep reported by Shih (1988) implies that there is a long-distance carryover effect in Mandarin, but Lin and Yan (1991) suggest that the scope of tonal interaction is limited to only adjacent syllables in this language. Laniran (1992) and Laniran and Clements (1995) found long-distance anticipatory effects in Yoruba, whereas Xu (1993) failed to find any consistent long-distance anticipatory effect in Mandarin. As for the nature of the contextual variations, although Gandour *et al.* (1992), Gandour *et al.* (1994) and Xu (1993, 1997) provided some speculation, many more details need to be learned before an adequate understanding of their mechanism can be attained. These details include the shape and height of the f_0 contours as well as their alignment with the syllables they are associated with.

The alignment of f_0 contours has been the concern of several recent studies (Steele, 1986; Silverman & Pierrehumbert, 1990; Prieto, Santen & Hirschberg, 1995; Arvaniti, Ladd & Mennen, 1998). So far, however, the findings have been diverse and further research is needed. One of the difficulties with the alignment research is again related to the issue of underlying specifications of the pitch targets under study. In the case of

non-tone languages, f_0 peaks are usually associated with pitch accents or stress. Since they do not usually occur in isolation, it is hard to know precisely what their underlying specifications are. For this reason, and for reasons stated earlier, the present study intends to investigate f_0 variations in Mandarin, a language with four contour tones (H, R, L and F, as described earlier), whose underlying specifications have been relatively well studied.

The effect of focus on the formation of f_0 contours will also be examined in the present study. Previous studies have found that focus contributes much to the overall f_0 pattern of an utterance. Pierrehumbert (1980) examined the relative f_0 heights of an early pitch accent and a later one in an utterance as a function of focus location. Her data suggest that when there is an early focus in the utterance, the f_0 range in the later portion of the utterance is reduced, whereas the earlier f_0 contour is only slightly lowered when the focus is on a later pitch accent. In a series of studies by Cooper and Eady and their colleagues (Cooper, Eady & Mueller, 1985; Eady & Cooper, 1986; Eady, Cooper, Klouda, Mueller & Lotts, 1986), it was found that the effect of a narrow focus in an English sentence is to raise the f_0 of the focused word and to lower the f_0 of the later words in the sentence. In contrast to the lowered f_0 of the post-focus words, however, f_0 of the pre-focus words was found to remain much the same as in a focus-neutral sentence. Gårding (1987) reported a similar asymmetry of f_0 variation around the focus for both Mandarin and other languages. More recently, Jin (1996) also looked at the effect of focus on the f_0 curves in Mandarin, and his data suggest similar asymmetry. These findings indicate that when occurring early in an utterance, a focus has the effect of tilting the overall f_0 curve downward over the course of the utterance. This effect is once again somewhat similar to the overall effect of declination. It may also overlap with the effect of downstep when both are present in an utterance. In addition, in a recent pilot study (Xu & Kim, 1996), it was noticed that subjects often voluntarily put a focus into a sentence they were reading aloud, sometimes to correct a mistake, sometimes to make a contrast to the previous sentence, and sometimes for no apparent reason. It is therefore possible that speakers always use some kind of focus pattern when saying a sentence. If they are not told explicitly which pattern to use (and one is not obvious, as in a carrier like "Say __ again"), speakers may simply pick one themselves. It thus seems advisable that focus be directly controlled by including different focus conditions in the design of the experiment.

The present study therefore investigates the formation of f_0 contours under an interactive influence of lexical tone and focus by examining short Mandarin sentences with systematically varied tonal components and sentence foci. In particular, the study tries to answer the following questions. First, how can tone and focus be implemented simultaneously when both have to use fundamental frequency as their major acoustic correlate? Second, how can lexical tones be adequately implemented acoustically, given that they may interfere with each other, as has been widely reported? Third, how do the interaction of tone and focus and the interaction among the tones themselves affect the shape of f_0 contours and their alignment with the syllabic elements of an utterance?

2. Methods

2.1. *Material*

The factors manipulated in the experiment were segmental composition of syllables, lexical tones, focus, and sentence type. Of these, only lexical tones and focus were under

TABLE I. Tone patterns and corresponding sentences used as recording material. H, R, L, and F represent high, rising, low, and falling tones, respectively

Word 1		Word 2		Word 3	
HH māomī̄	'kitty'	H mō	'touches'	HH māomī̄	'kitty'
HR māomí	'cat-fan'	R ná	'takes'	LH mǎdāo	'sabre'
HL māomī̀	'cat-rice'	F mài	'sells'		
HF māomì	'cat-honey'				

direct scrutiny. The control of the other two factors was done by keeping their variation to a minimum. To control for segmental effects, simple CV syllables with a sonorant as the initial consonant (except for one, due to difficulty in finding a real word) were used. The effects of initial consonants on the f_0 of a syllable have been well established (Lehiste & Peterson, 1961; Lehiste, 1970; Howie, 1974; Hombert, 1978; Rose, 1998; Santen & Hirschberg, 1994), and sonorants are known to present the least disturbance and interruption of the continuity of f_0 contours. Hence, they were used as much as possible in the experiment. For ease of segmentation, the voiced sonorants used were all nasals. Due to the abrupt shift of resonance cavities upon nasal closure and release, vowel-nasal boundaries can be seen clearly both in the waveform and in the spectrogram, as will be illustrated later. For male speakers, the shift usually occurs within two adjacent vocal cycles; for female speakers, within two to three vocal cycles. To control for the effect of sentence type, only declaratives were used, because they are known to show the most f_0 declination.

For the factors under direct scrutiny, systematic manipulations were administered. The sentences used in the experiment consisted of three words (two disyllabic and one monosyllabic) as shown in Table I. To control for the effect of lexical tones, the second, third, and fourth syllables in these utterances had varying tones. The second syllable had four alternating tones: H, R, L, and F. The third syllable had three alternating tones: H, R, and F. The fourth syllable had two alternating tones: H and L. The L tone was not used on the third syllable so as to avoid the phonological tone sandhi that would change a LL sequence into a RL sequence (Wang & Li, 1967; Chao, 1968). Only two tones were used on the fourth syllable in order to reduce the amount of data.

To control for the effect of focus, each sentence was preceded by a question asking about a specific piece of information in the sentence. For example, for the sentence "māomī̄ mō māomī̄" four different precursor questions were used, as shown in Table II. Four focus conditions were thus created: (a) neutral focus, (b) focus on word 1, (c) focus on word 2, and (d) focus on word 3.

The target sentences and their precursor questions were repeated five times, randomized, and printed in Chinese. The total number of sentence pairs was

$$4\ (\text{1st word}) \times 3\ (\text{2nd word}) \times 2\ (\text{3rd word}) \times 4\ (\text{focus location}) \times 5\ (\text{repetition}) = 480.$$

A pilot study showed that some subjects experienced difficulty when the tone of second syllable varied in the same reading list. They often made mistakes because of the phonetic similarities among the sentences, and they sometimes put focus in the wrong place in an effort to prevent such mistakes. To alleviate this problem, the sentences were divided into four groups, each consisting of 120 sentences. In each group

TABLE II. Questions preceding the target sentences

1. Māomī gànmá ne?	(What is kitty doing?)
2. Shéi mō māomī?	(Who is touching kitty?)
3. Māomī zěnmo nòng māomī?	(What is kitty doing to kitty?)
4. Māomī mō shénmo?	(What is kitty touching?)

the first (disyllabic) word in all sentences was kept constant, while the rest of the words varied.

On the printed lists, the words to receive focus were underscored to remind the subjects of the location of the focus. (No words were underscored in sentences preceded by question 1 in Table II.) The precursor questions made the desired foci natural, while the underscores helped to reduce the number of errors.

2.2. *Subjects*

Studies of f_0 contours typically examine data from only a few speakers (usually ranging from 1 to 4). The advantages of such an approach is that data from each individual speaker can be displayed and discussed in detail in published reports, and individual variations can be described. There are two main disadvantages to this approach. First, it may become very cumbersome unless the number of subjects is very small. Second, the analysis is really a series of case studies of the specific individuals, and there is no statistical basis for generalizing to a larger population of speakers. An alternative approach is to collect data from more than two or three speakers. Such an approach was adopted by Xu (1997), using 8 speakers. This makes it possible to conduct repeated measures statistical analyses that test the reliability of effects across subjects. To the extent that the subjects can be viewed as a random sample of a larger population, legitimate statistical inference can be drawn about that larger population.

Eight native speakers of Mandarin (including the author), 4 males and 4 females, recorded the sentences. Six were graduate students studying at Northwestern University, and one was a school teacher in Beijing before coming to the US. Except for the author, they were all born and raised in Beijing, China. The author is a native speaker of Standard Chinese (Putonghua), which closely resembles Beijing Mandarin phonetically.

2.3. *Recording*

Recording was conducted in a sound treated booth in the Department of Communication Sciences and Disorders, Northwestern University. A condenser microphone was placed about 12 inches in front of the subject's mouth. Subjects were instructed to read aloud both the precursor questions and the target sentences, putting focus on the underscored words when reading the target sentences. For sentences preceded by question 1 in Table II, they were told not to emphasize any word. During recording, when the experimenter determined that a particular sentence was not produced properly, the subject was asked to repeat both the question and the sentence. Each subject read the sentences in four sessions, with a 2–5-minute break in between. In each session the subject read 120 sentences in which the first word was the same, as described earlier. The

order of the sessions was different for each subject. The speech signals were directly digitized onto the hard disk of a Macintosh 7500/100 (with built-in 16 bit sound) at a sampling rate of 22 kHz, using SoundEdit, a software digitization program by Macromedia.

2.4. *f_0 and timing measurements*

The digitized signals were transferred to a Sun Sparc 5 workstation and analyzed by the ESPS signal processing software package (Entropic Inc). The ESPS *epochs* program was used to mark every vocal cycle in the sentences. After screening, additional hand-editing was done, as necessary, to correct spurious vocal pulse labeling by the *epochs* program (such as double-marking or vocal-cycle skipping). The marked sentences were then manually labeled in the ESPS *xwaves* program for the onset and offset of each consonant and vowel segment using the *xlabel* program. An example of a marked waveform is shown in Fig. 1. As can be seen, a V-N boundary is located at the vocal pulse where the oral cavity is determined to be closed; a N-V boundary is located at the vocal pulse where the oral cavity is determined to be open. Evidence for the moments of oral cavity opening and closing was derived both from the spectrogram and from the waveform, as can be seen in the figure.

The vocal pulse markings and segment labels for each utterance were saved and subsequently processed by a set of computer programs written by the author. In addition to other specialized computations, each of these programs converted the duration of vocal cycles into f_0 values, and smoothed the resulting f_0 curve using the *trimming algorithm* described in Appendix 1.

The trimming algorithm was particularly effective in smoothing out sharp spikes in the raw f_0 tracing often seen around nasal-vowel junctions, such as those shown by the thin line in Fig. 2. Sharp spikes also often occur when the vocal-cycle-marking program shifts its marking from one of the multiple peaks or valleys in a vocal cycle to the other, as happened in the vowel segment of /na/ and the second /mao/ in Fig. 2. The algorithm effectively trims out these sharp spikes, as shown by the thick line in Fig. 2.

To visually inspect and compare the f_0 contours, the smoothed f_0 curves were further processed to (a) time-normalize each f_0 curve for every consonant and vowel segment, i.e., taking a predetermined number of f_0 points at equal time intervals from the smoothed f_0 curve of each segment, and (b) average over the 5 repetitions of the same sentence in a particular focus condition. The resulting smoothed, time-normalized, and averaged f_0 curves were saved into a single file for each subject. For displays used in this paper, the f_0 values obtained from each subject were first converted to their logarithms (in order to accommodate the pitch range difference among speakers, especially between the male and female speakers). These logarithmic values were then averaged across subjects. Then the averaged logarithmic values were converted back to f_0 values in Hz, which were used for display and visual inspection.

The segmentation provided accurate tone-segment alignment information. The smoothing not only reduced random variation in the f_0 contours, but also assured the subsequent accurate measurement of the location and value of f_0 peaks and valleys. The time normalization served two major purposes. First, it made averaging across the repetitions of the same sentence possible. Second, it facilitated direct comparison among different f_0 curves. Time normalization did throw away part of the durational information that is also relevant for intonation, but that was done only for the graphic

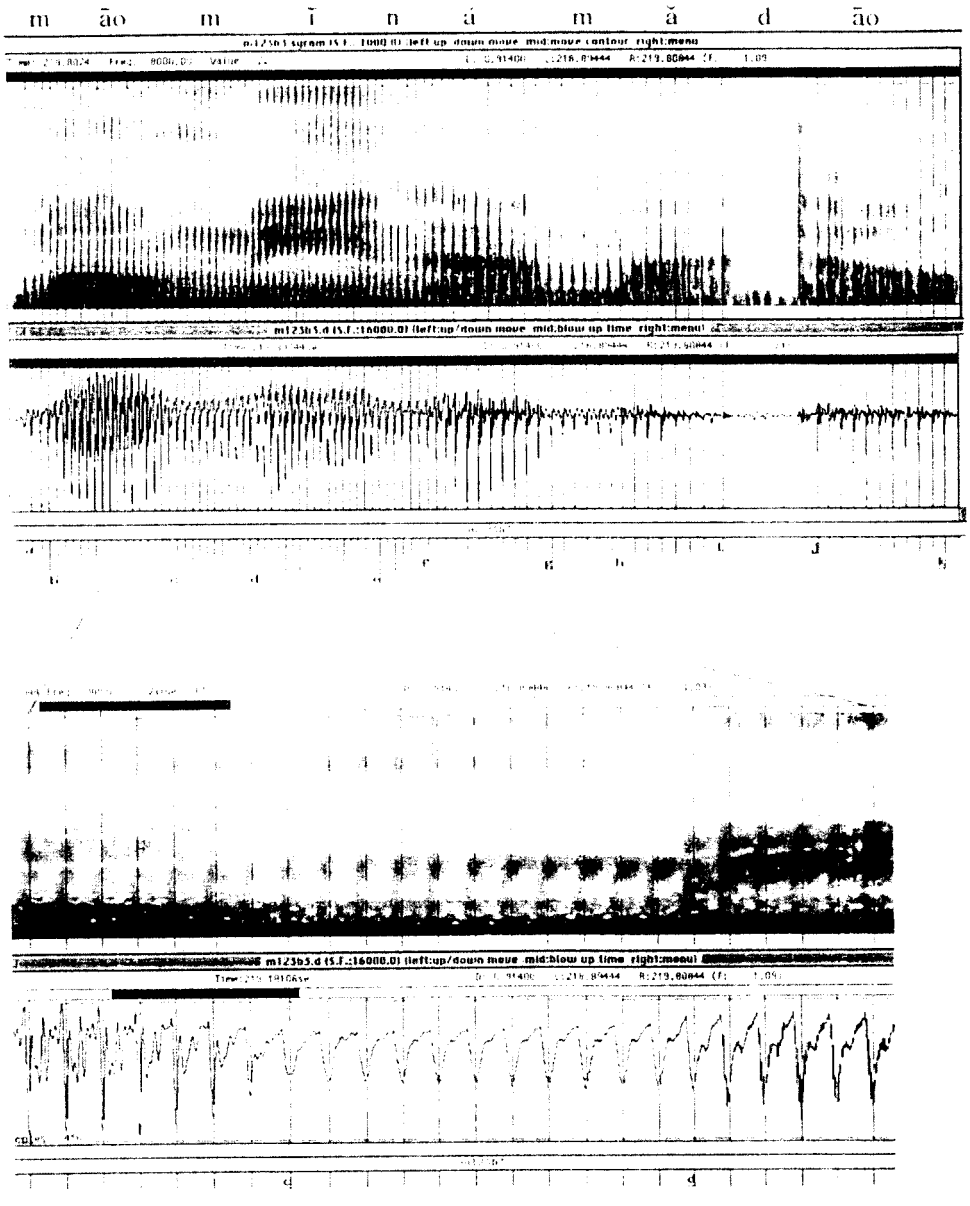

Figure 1. Top panel: an example of consonant and vowel segmentation and vocal-cycle marking. The vocal-cycle marking was produced by the "*epochs*" program in the ESPS package. The segmentation was done manually. A V–N boundary is located in the vocal pulse where the oral cavity is evidently closed; a N–V boundary is located in the vocal pulse where the oral cavity is evidently open. Evidence for the instants of oral cavity opening and closing could be found both in the spectrogram and in the waveform, as shown in the expanded view of the spectrogram and waveform in the bottom panel.

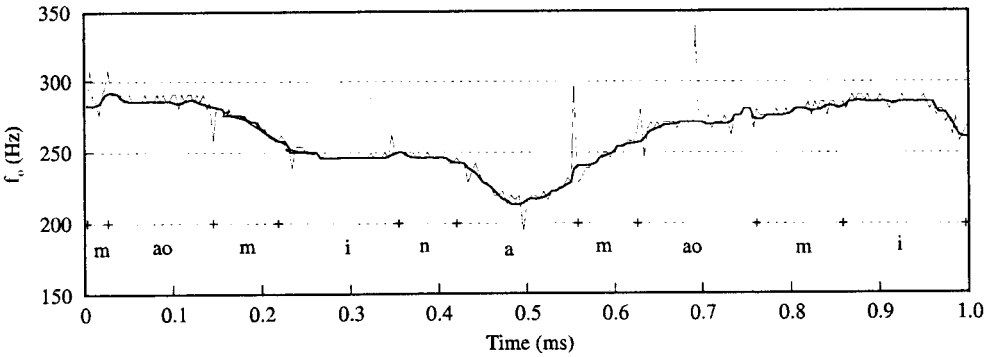

Figure 2. A raw f_0 curve obtained by taking the inverse of vocal periods (thin line), and the same curve after being smoothed by the trimming algorithm described in Appendix 1.

displays. Since the timing of all the vocal cycles and all the segment boundaries in each utterance was recorded, no duration information was lost in the data and durations were actually used extensively in some of the analyses.

3. Analysis and results

The goal of the f_0 analysis is to determine the contribution of lexical tone and focus to the formation of f_0 curves and to find out how much and what kind of contribution they each make. The first step is to find a way to graphically display the f_0 curves so that the effects of contributing factors can be visually examined. Time-normalization within each segment as described above made it possible to average across repetitions of the same sentence produced by each speaker (and further across different speakers for the displays in the present paper). The averaging reduced random utterance-to-utterance fluctuation, letting those variations most consistent across repetitions stand out. These variations could then be more closely analyzed statistically to evaluate their magnitude and consistency.

The following few figures (Figs. 3–6) display mean f_0 curves averaged across all eight speakers and across all five repetitions of the same sentence produced with the same focus pattern. These curves are displayed in such a way that the effects of focus and tone are readily visible. This is done by overlaying in each panel f_0 curves that differ in only one of the conditions: focus of the sentence (Fig. 3), tone of the first word (Fig. 4), tone of the second word (Fig. 5), or tone of the third word (Fig. 6). During data analysis, separate graphs were made for each subject. Since it is impossible to show all of them efficiently, only the averaged curves are displayed. The statistical analyses described below were conducted to ascertain the reliability of any apparently systematic variation observed in the graphical displays.

3.1. *Effect of focus*

3.1.1. f_0

Figure 3 shows the effect of focus. In each panel of Fig. 3, mean f_0 curves of the same sentences produced in the four focus conditions are displayed: neutral focus (neutral),

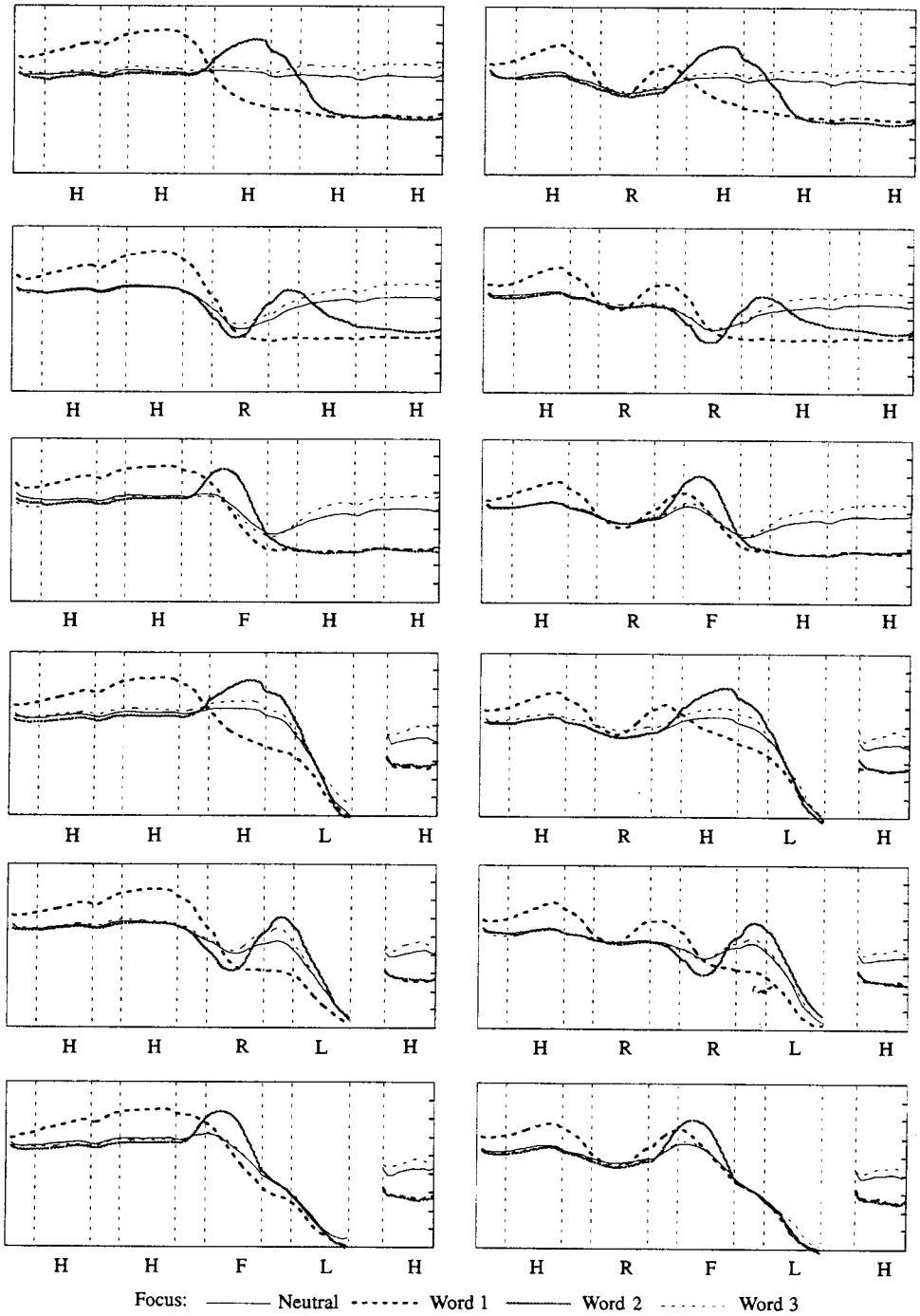

Figure 3. Effects of focus on f_0 curves. In each panel, the tonal composition is held constant, while the focus varies among neutral, word 1, word 2, and word 3. Individual panels are referred to in the discussion using their column and row index. Thus, C1R1 refers to the panel in the top row of the leftmost column.

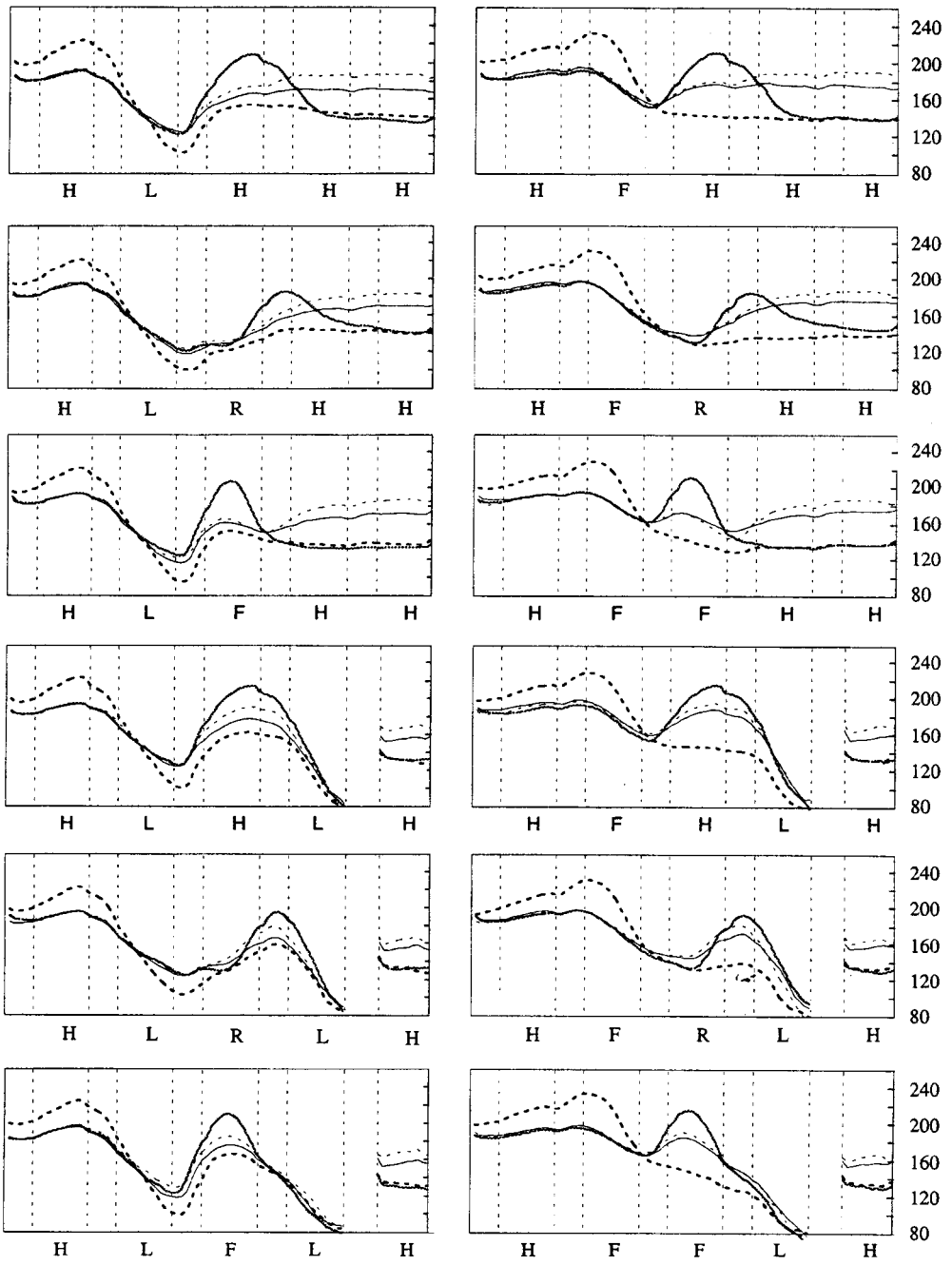

Figure 3. *Continued.*

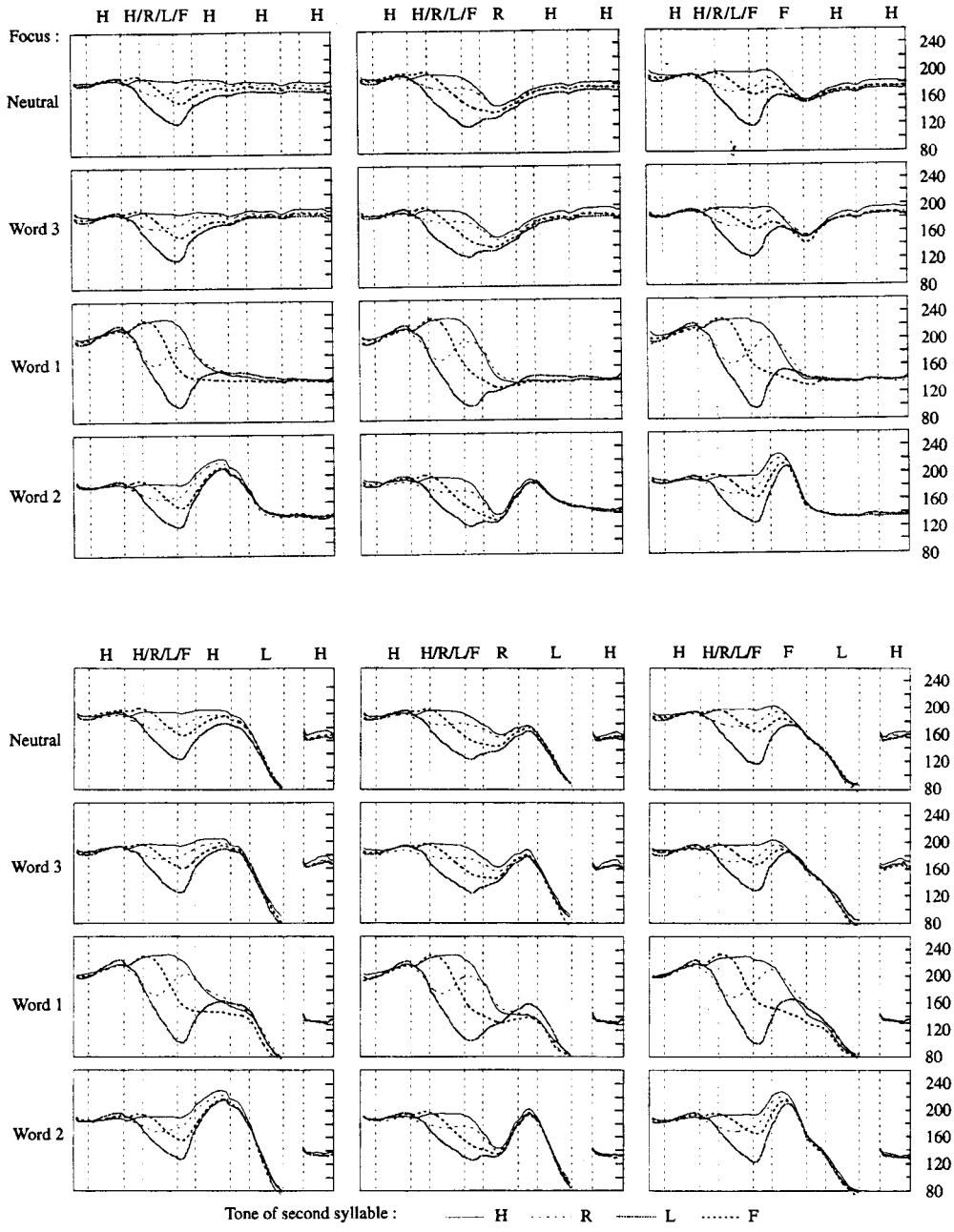

Figure 4. Effects of the tone of the second syllable on f_0 curves of the entire utterance. In each panel, the tone of the second syllable varies from H, R, L, to F, while the tones of all other syllables are held constant. Individual panels are referred to in the discussion using their column and row index. Thus, C1R1 refers to the panel in the top row of the leftmost column.

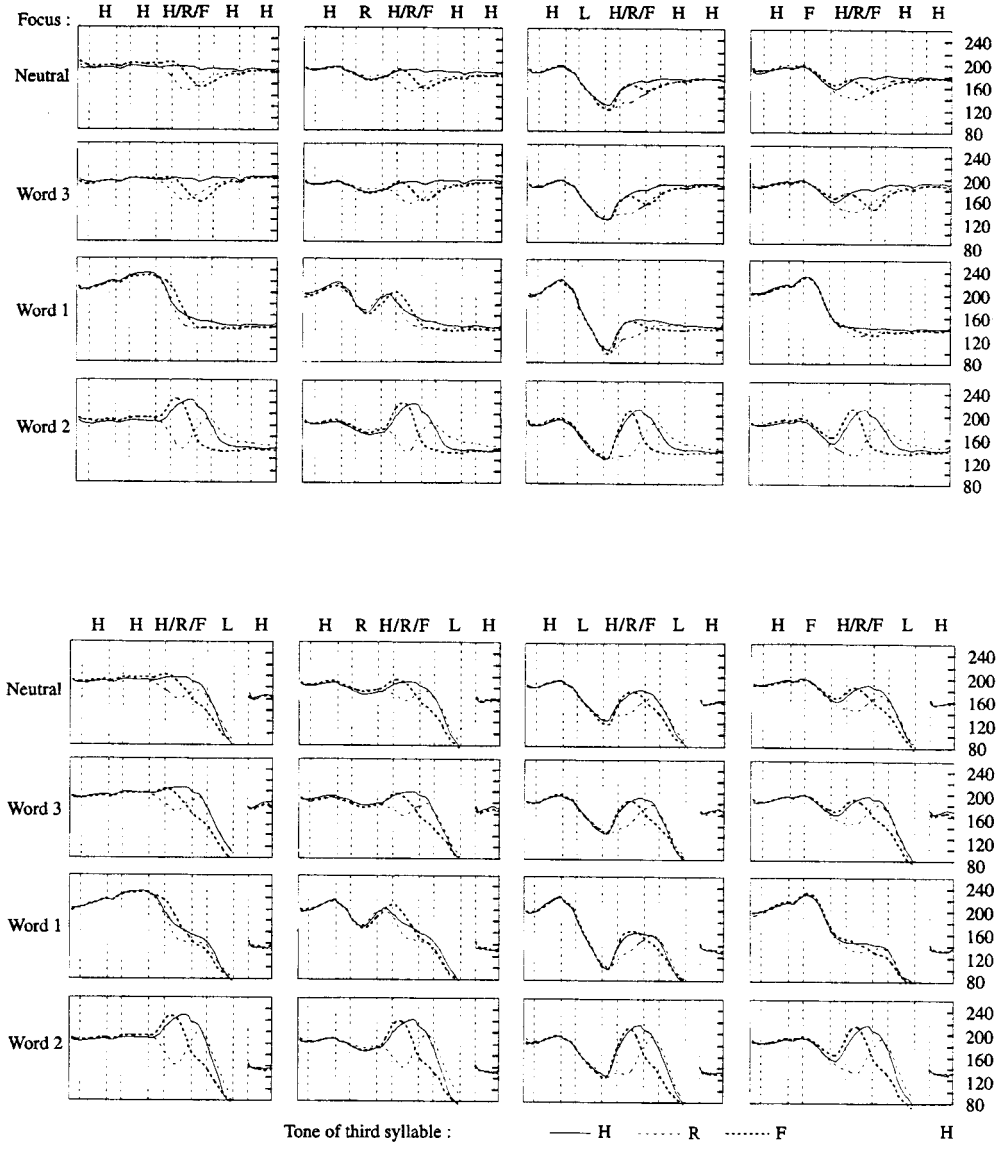

Figure 5. Effects of the tone of the third syllable on f_0 curves of the entire utterance. In each panel, the tone of the third syllable varies from H, R, to F, while the tones of all other syllables are held constant. Individual panels are referred to in the discussion using their column and row index. Thus, C1R1 refers to the panel in the top row of the leftmost column.

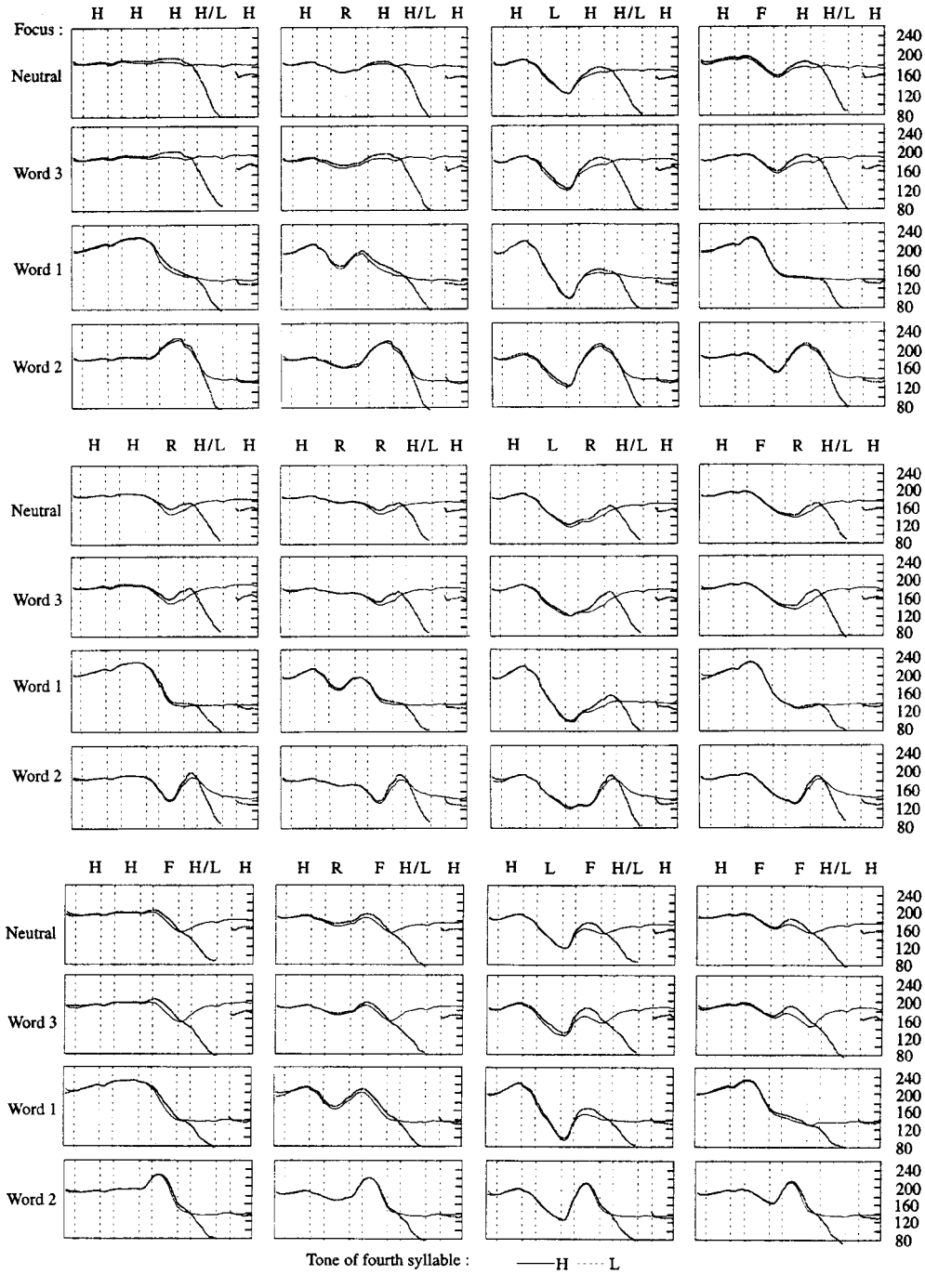

Figure 6. Effects of the tone of the fourth syllable on f_0 curves of the entire utterance. In each panel, the tone of the fourth syllable varies between H and L, while the tones of all other syllables are held constant. Individual panels are referred to in the discussion using their column and row index. Thus, C1R1 refers to the panel in the top row of the leftmost column.

focus on the first word (word 1), focus on the second word (word 2), and focus on the third word (word 3). In most panels, substantial variations among different focus conditions can be seen. For example, in panel C1 R1 (Column 1, Row 1), the four sentences all consist of the H tone only. However, the height and shape of the f_0 curves differ extensively due to different focus conditions. Compared to the neutral-focus curve, which is virtually flat, the f_0 of the first two syllables is raised and that of the following syllables lowered when the focus is on the first word. When the focus is on the second word, which is monosyllabic, the f_0 of that word is raised and that of the following word lowered. Since Mandarin has lexical contour tones (which conceivably might have left relatively little room for f_0 variation), such substantial f_0 differences due to focus alone are quite remarkable. Interestingly, when the focus is on the last word, although the f_0 of that word is raised, the magnitude of the rise is much smaller compared to that caused by an earlier focus in the sentence.

Closer examination of Fig. 3 shows that the patterns observed in the H-tone-only sentence just described reflect the general effects of focus on f_0 contours. First, under focus, the high points of the H, R, and F tones are raised and the low points in the R, F, and L tones are lowered. In other words, it seems that the f_0 range is expanded by focus: high points become higher and low points become lower. In the case of R (panel in R2 and R5 and in C2) and F (panels in R3 and R6 and in C4) tones, both their high points are raised and low points lowered, although the magnitude of the raising is greater than that of the lowering.[1] Secondly, as can be seen in all panels, the f_0 of all the words *after* the focus is substantially lowered, regardless of whether the f_0 of the focused word has been raised or lowered (due to the f_0 range expansion). Neither of these two effects, however, can be clearly observed when the focus is on the last word. Although the f_0 of the last syllable is higher than that of the neutral-focus condition, the increase seems much smaller than when the focus is earlier in the sentence. In addition, since there are no post-focus words, no post-focus lowering can be observed. As a result, there is little change in the overall shape of the f_0 contour as compared to the neutral focus condition. Finally, in contrast to the substantial on-focus raising and post-focus lowering, the f_0 of the pre-focus words is barely changed. In the case of word 1, in particular, the f_0 contours in the neutral-focus and pre-focus conditions virtually coincide in all the panels.

To further examine these patterns, maximum and minimum f_0 values were measured for different focus conditions in three of the syllables: the second syllable of word 1, which carries 4 different tones (H, R, L, F), word 2, which is monosyllabic and carries 3 different tones (H, R, F) and the first syllable of word 3, which carries 2 different tones (H, L). The focus conditions for each word are *neutral* — no narrow focus in the entire utterance, *pre-focus* — when the word occurs before focus, *focus* — when the word is under focus, and *post-focus* — when the word occurs after focus.[2] The mean maximum and minimum f_0 values (averaged across eight subjects) are displayed in the left four columns in Table III. The right six columns display differences between the mean f_0 values in the left four columns. The difference in each cell is calculated by subtracting the mean f_0 value in one condition from another, as indicated by the column heading. To test the significance of these differences, two-tailed paired *t* tests were conducted, in which subject is treated as

[1] The greatest high-raising in the R tone and low-lowering in the F tone are observed in the initial nasal following the focused word due to the carryover effect to be discussed later.

[2] Due to the carryover effects to be discussed later, the maximum or minimum f_0 of a word sometimes occurred in the initial nasal of the following word.

the random factor. To adjust for potential significance inflation due to multiple comparisons, beside the commonly-used probability levels of 0.05 and 0.01, a third level of significance was computed using the Bonferroni adjustement: $p = 0.05/54 = 0.00093$.

As can be seen in Table III, in all three words, the maximum f_0 is higher in the focus condition than in the neutral focus condition., The mean differences range from 18 to 39 Hz. In word 1, the maximum f_0 is not significantly different between the pre-focus and neutral focus conditions. In word 2, the maximum f_0 is not different between pre-focus and neutral focus conditions for the F tone, and only marginally different for the H and R tones. In the latter case, the maximum f_0 is higher (rather than lower) in the pre-focus condition than in the neutral focus condition. Also in word 2, the maximum f_0 in the post-focus condition is significantly lower than in the pre-focus condition for all three tones (marginally for the F tone), and lower (with marginal significance) than in the neutral focus condition for the H tone. In word 3, the maximum f_0 is significantly lower in the post-focus condition than in the neutral focus condition (by 27 Hz in the H tone).

The minimum f_0 in word 1 under focus is significantly lower than in pre-focus and neutral focus conditions for the L and F tones (ranging from 9 to 29 Hz), but not for the R tone. In word 2, the minimum f_0 is lower under focus than in pre-focus and neutral focus conditions for the F and R tones, but higher than in post-focus condition for the F tone. Also in word 2, the minimum f_0 in post-focus condition is lower than in both pre-focus and neutral focus conditions. The minimum f_0 in word 3 is not significantly different across the focus conditions.

Table IV displays f_0 ranges in different focus conditions for the three word positions (left four columns), and the differences between them (right six columns). Each f_0 range is computed by subtracting the lowest f_0 from the highest f_0 for a word position in a particular focus condition. Two-tailed paired *t* tests were conducted to test the significance of these differences, the results of which are indicated by the superscripts in the table. As can be seen, the f_0 ranges in the focus condition are greater than in all other conditions (by as much as 66 Hz), while the f_0 ranges in the post-focus condition is reduced from that in the neutral focus condition (by as much as 20 Hz in word 3). The only other substantial difference is between the neutral focus and post-focus conditions for word 3, indicating a significant reduction in the f_0 range when the word is post focus. Note that the f_0 range variations shown in Table IV do not demonstrate the full scope of f_0 range changes due to focus, because word 2 and word 3 do not have all the four tones. Nevertheless, the differences shown in the table are still fairly substantial. In contrast, the difference in f_0 range between the neutral focus and pre-focus conditions is quite small.

In general, therefore, there seems to be a radical *asymmetry* around the focus: the f_0 range at the focus is substantially expanded; the f_0 range after the focus is lowered as well as compressed; and the f_0 range before the focus does not really deviate much from the neutral focus condition. In other words, there appear to be three distinct focus-related pitch ranges: *expanded* in non-final focused words, *suppressed* (lowered and compressed) in post-focus words, and *neutral* in all other words.

3.1.2. *Duration*

Duration of the five syllables in all the sentences was measured for different focus conditions. The mean duration measurements are shown in the left half of Table V. The right half of the table displays results of two-tailed *t* tests comparing duration of different focus conditions. As shown in the table, syllable duration increases significantly under

TABLE III. Focus effect. Left four columns: mean maximum and minimum f_0 (averaged across eight speakers) of word 1, 2, and 3 in four conditions: neutral focus (neutral), on-focus (focus), pre-focus (pre), and post-focus (post). For each word, the maximum and minimum f_0 values are located in the vowel(s) of the word and the initial nasal of the following word. Right six columns: mean differences between f_0 values in the left four columns. In each cell, the difference is calculated by subtracting the mean maximum or minimum f_0 of the second focus condition from the first as indicated by the column heading

	neutral	pre	focus	post	focus − neutral	focus − pre	focus − post	pre − neutral	pre − post	neutral − post
Word 1										
H-tone max	216	220	251		35[c]	32[b]		3		
R-tone max	199	205	217		18[a]	12[b]		5		
F-tone max	210	208	249		39[c]	41[b]		− 1		
R-tone min	180	177	177		− 2	1		− 3		
L-tone min	129	131	102		− 27[a]	− 30[a]		3		
F-tone min	167	164	155		− 12[a]	− 9[b]		− 3		
Word 2										
H-tone max	202	211	234	182	32[c]	24[c]	52[c]	9[a]	29[b]	20[a]
R-tone max	185	194	208	177	24[b]	14[c]	31[b]	10[a]	17[b]	8
F-tone max	201	203	235	194	35[c]	32[c]	41[c]	2	9[a]	6
R-tone min	153	155	141	138	− 12[a]	− 14[b]	4	2	17[b]	16[b]
F-tone min	156	155	146	140	− 11[a]	− 10[b]	6[b]	− 1	16[b]	16[a]
Word 3										
H-tone max	194	208	167		14[a]		41[c]			27[b]
L-tone min	104	104	97		− 1		7			7

Note: The superscripts a, b, and c indicate probability values of $p < 0.05$, $p < 0.01$ and $p < 0.05/54 = 0.00093$ (Bonferroni adjustment for multiple comparisons) in two-tailed paired t tests, respectively. In the t tests, $df = 7$ (i.e. each subject contributes one difference score to each test).

TABLE IV. f_0 ranges in different focus locations and their comparisons. The values in the left four columns are computed by subtracting the lowest f_0 from the highest f_0 in each word position in a particular focus condition. The values in the right 6 columns are differences among f_0 ranges displayed in the left four columns, as indicated by the column headings

	neutral	pre-focus	focus	post-focus	focus − neutral	focus − pre	focus − post	neut − pre	neut − post	pre − post
Word 1	89	90	155		66[c]	65[c]		0		
Word 2	51	58	97	59	46[c]	39[c]	38[c]	− 7[a]	− 8	− 1
Word 3	90		104	70	15[b]		34[c]		20[c]	

Note. The superscripts a, b, and c indicate probability values of $p < 0.05$, $p < 0.01$ and $p < 0.05/12 = 0.0042$ (Bonferroni adjustment for multiple comparisons) in two-tailed paired t tests, respectively. In all t tests, $df = 7$.

TABLE V. Left four columns: mean syllable duration in ms across all speakers in four focus conditions: neutral focus (neutral), on-focus (focus), pre-focus (pre), and post-focus (post). Right six columns: mean differences between the duration values in the left four columns. In each cell, the difference is calculated by subtracting the mean duration in the second focus condition from the first as indicated by the column heading

	neutral	pre	focus	post	focus − neut	focus − pre	focus − post	pre − neut	pre − post	neut − post
syllable 1	125	123	148		23[c]	25[c]		− 2		
syllable 2	161	163	183		21[b]	20[c]		2		
syllable 3	181	183	233	178	52[c]	50[c]	55[c]	2	5	3
syllable 4	190		202	190	12[b]		12[a]			0
syllable 5	245		258	229	13[b]		28[c]			16[a]

Note. The superscripts, a, b, and c indicate probability values of $p < 0.05$, $p < 0.01$ and $p < 0.05/18 = 0.0028$ (Bonferroni adjustment for multiple comparisons) in two-tailed paired t tests, respectively. In all t tests, $df = 7$.

focus, regardless of the position of the syllable in the utterance. However, duration does not differ significantly between the neutral and pre-focus conditions for syllables 1–3 (word 1 and 2). Nor does it differ significantly between neutral and post-focus conditions in syllable 3 (word 2). It is interesting that the amount of increase in duration seems more comparable between words than between syllables. The combined duration increase due to focus is around 50–55 ms for both word 1 (disyllabic) and word 2 (monosyllabic), and it is 25–41 ms for word 3, whereas the increase in individual syllables in the disyllabic word is much smaller than the increase in the monosyllabic word.

3.2. *Effect of tone*

As discussed in the Introduction, it has been well established that the major acoustic correlate of lexical tones in Mandarin is fundamental frequency, and the basic f_0 contours of these tones produced both in isolation and in context have been extensively studied. At issue here is how much lexical tones contribute to the local and global f_0 contours. Figs. 4–6 display all the mean f_0 curves in a manner that allows easy visual inspection of the f_0 variations due to the effect of tone of the second, third, and fourth syllables in the utterances. In these figures, all the tones in each panel remain constant except for the tone of the second (Fig. 4), third (Fig. 5), or fourth syllable (Fig. 6), respectively.

It can be seen in Figs. 4–6 that, although the greatest f_0 variations occur during the syllables that carry the varying tones, f_0 contours in the following syllables also vary substantially, especially during the initial nasal consonants. In contrast, f_0 contours in the preceding syllables vary little with the following tones. This basic pattern agrees with findings reported by previous studies (Xu, 1993, 1997), except that the dissimilatory anticipatory influence reported in those studies is not quite so obvious here. Only in Fig. 6 can the kind of anticipatory effect reported previously be seen clearly. In most of the graphs in Fig. 6, the H tone on syllable 3 has a higher f_0 when followed by the L tone than by the H tone.

To examine these variations more closely, average maximum and minimum f_0 values of all five syllables (measured in the vowel portion of each syllable) in all sentences were obtained. Table 6 displays the mean maximum and minimum f_0 averaged across all eight subjects. Each panel in Table 6 displays the effect of the tones of the 2nd, 3rd, or 4th syllable in the utterances. For each syllable, the focus conditions are indicated by the column heading: Neut for neutral focus and W 1–3 for focus on words 1–3, respectively.

These measurements were used as dependent variables in a series of three-factor repeated-measure ANOVAs. The independent variable used were (1) tone of the second syllable (H, R, L, F), (2) tone of the third syllable (H, R, F), and (3) tone of the fourth syllable (H, L). A separate ANOVA was conducted for each of the four focus conditions. Significant probability values of the ANOVAs are shown at the bottom of the columns.

3.2.1. *Effect of tone proper*

As shown in Table VI, variations in the maximum and minimum f_0 in syllables 2, 3, and 4 due to the effect of tones carried by these syllables are all highly significant (in both Table VIa and Table VIb: columns 5–8, top panel, columns 9–12, middle panel, and columns 13–16, bottom panel, with the only exception of maximum f_0 in syllable 4 when focus is on word 2 (focus = word 2, bottom panel of Table VIa). A close examination of

TABLE VI. Mean maximum (a) and minimum (b) f_0 values of syllables 1–5 measured in the vowel of each syllable. Each panel displays the effect of the tones of the 2nd, 3rd, or 4th syllable in the utterances. For each syllable, the focus conditions are indicated by the column heading. Neut for neutral focus, W 1-3 for focus on word 1–3 respectively. Significant probability values of 3-factors repeated measure ANOVAs are shown at the bottom of the columns

		Syllable 1				Syllable 2				Syllable 3				Syllable 4				Syllable 5			
Focus condition:		Neut	W1	W2	W3	Neut	W1	W2	W3	Neut	W1	W2	W3	Neut	W1	W2	W3	Neut	W1	W2	W3
(a)																					
	H	211	237	206	209	214	251	209	214	206	208	228	208	184	150	181	193	195	156	156	208
Tone of	R	207	234	204	206	191	208	188	192	198	203	220	201	179	150	175	186	187	155	155	199
syllable 2	L	210	240	209	210	175	187	176	177	177	165	212	187	172	156	176	184	182	156	153	198
	F	211	235	207	209	210	249	206	208	188	157	215	192	179	145	177	187	187	154	154	197
$P =$						0.001	0.001	0.001	0.001	0.001	0.001	0.012	0.001	0.011	0.003			0.008			0.049
Tone of	H	209	238	205	209	196	224	192	198	201	182	234	208	186	156	188	195	188	156	154	203
syllable 3	R	210	236	208	209	198	225	197	198	175	173	188	181	182	151	199	192	186	155	157	199
	F	210	236	207	208	198	222	196	198	200	194	235	203	167	144	144	174	188	155	153	200
$P =$		0.049				0.001	0.028	0.006		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001			0.002	0.002
Tone of	H	209	237	206	208	196	223	194	196	187	179	216	189	192	154	178	205	197	158	158	212
syllable 4	L	210	237	207	209	199	225	195	200	197	187	221	205	165	146	176	170	178	152	151	190
$P =$					0.018	0.035			0.005	0.005	0.001	0.003	0.003	0.001	0.002		0.001	0.002	0.007	0.001	0.002
(b)																					
	H	201	224	196	200	208	241	202	207	180	156	180	182	146	121	125	151	167	134	135	175
Tone of	R	195	216	192	195	182	177	178	183	176	160	174	178	142	122	125	148	160	135	133	170
syllable 2	L	196	217	195	195	132	108	133	136	156	141	162	161	140	125	123	145	158	134	132	168
	F	200	217	195	198	174	176	168	174	167	143	166	167	144	119	126	144	163	134	135	167
$P =$						0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.011				0.013			
Tone of	H	198	220	193	197	173	176	168	174	188	159	203	192	147	123	122	153	163	135	131	172
syllable 3	R	198	218	196	197	174	176	171	175	154	140	141	156	145	123	136	148	162	134	139	170
	F	199	218	195	197	175	174	172	176	168	151	168	168	137	118	116	140	161	133	131	169
$P =$			0.048					0.009		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001			0.001	
Tone of	H	198	219	194	196	172	174	169	173	167	147	168	167	181	147	152	190	186	147	147	200
syllable 4	L	199	218	195	197	175	177	172	177	173	153	173	177	104	96	98	104	138	121	121	141
$P =$								0.013	0.002	0.019	0.003	0.001	0.006	0.002	0.004	0.004	0.001	0.001	0.002	0.002	0.001

Fig. 6 indicates that the maximum f_0 in the L tone in syllable 4 mostly reflects the ending f_0 of syllable 3. When syllable 3 (word 2) is under focus, the maximum f_0 in the following L tone is raised so much that it is not very different from the maximum f_0 in the H tone. That, however, does not mean the contrast between the L and H tones is lost in this condition, because there is actually a large significant difference in the minimum f_0 between the two tones, as can be seen in column 15 in the bottom panel of Table VIb ($152 - 98 = 54$ Hz).

The f_0 variation due to tone increases when the tone is part of a word that is under focus. This is evident in Figs. 4–6 and in Table VI. For further verification, 4-factor repeated-measure ANOVAs were performed with maximum f_0 and minimum f_0 as dependent variables, and focus, tone of syllable 2, tone of syllable 3 and tone of syllable 4 as independent variables. For both maximum and minimum f_0, there are highly significant interactions between focus and tone proper. (For syllable 2, $F(9,63) = 10.00$, $p < 0.001$, $F(9,63) = 14.54$, $p < 0.001$, for $\max f_0$ and $\min f_0$, respectively; for syllable 3, $F(6,42) = 23.54$, $p < 0.001$, $F(6,42) = 23.71$, $p < 0.001$, for $\max f_0$ and $\min f_0$, respectively; for syllable 4, $F(3,21) = 33.95$, $p < 0.001$, $F(3,21) = 29.97$, $p < 0.001$, for $\max f_0$ and $\min f_0$, respectively.) These statistics indicate that when a word is under focus, maximum f_0 increases in all tones except for the L tone in syllable 4 (which is lower than when focus is on word 2), and minimum f_0 decreases in the R and L tones when they occur early in the sentence (syllables 2 and 3), but not later in the sentence (syllable 4).

When the other three variables are controlled for (i.e., averaged across all other conditions), variation in maximum f_0 due to tone proper is as much as 43 Hz in syllable 2 (between H and L tones), 29 Hz in syllable 3 (between H and R tones), and 18 Hz in syllable 4 (between H and L tones). Variation in minimum f_0 due to tone proper is as much as 84 Hz in syllable 2 (between H and L tones), 38 Hz in syllable 3 (between H and R tones), and 67 Hz in syllable 4 (between H and L tones). When all the tone variables are controlled for, variation in maximum f_0 due to focus is as much as 29 Hz in syllable 2 (between foci on word 1 and 2), 36 Hz in syllable 3 (between foci on word 1 and 2), and 37 Hz in syllable 4 (between foci on word 1 and 3); variation in minimum f_0 due to focus is as much as 5 Hz in syllable 2 (between foci on word 1 and 2), 22 Hz in syllable 3 (between foci on word 1 and 3), and 26 Hz in syllable 4 (between foci on word 1 and 3). In terms of magnitude, therefore, f_0 variations due to tone proper are greater than those due to focus, although the latter are also quite extensive. In addition, focus seems to have greater influence on maximum than on minimum f_0, where tone proper has greater influence on minimum than on maximum f_0.

3.2.2. *Carryover effects*

Table VI also reveals that a tone exerts strong carryover influence on the f_0 contour of the syllable that immediately follows it. As shown in columns 9–12 of the top panel, columns 13–16 of the middle panel, and columns 17–20 of the bottom panel in Table VIa and Table VIb, the immediate carryover effects are significant in all focus conditions. The largest f_0 differences occur when the influencing tone is under focus: 51 Hz in syllable 3 when focus is on word 1, 55 Hz in syllable 4 when focus is on word 2, and 22 Hz in syllable 5 when focus is on word 3. The much smaller difference in syllable 5 than in the other two syllables may be due to the fact that the syllable itself is part of the word that is under focus. The immediate carryover influence is always assimilatory: a high f_0 offset in the preceding tone (as in a H or R tone) raises the maximum f_0 of the following tone,

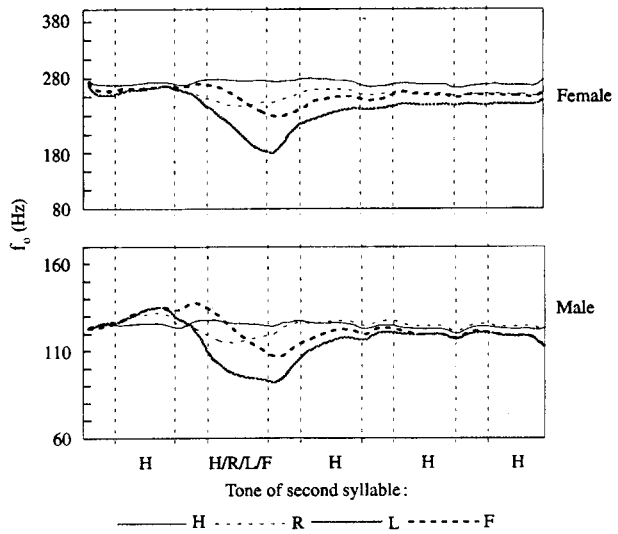

Figure 7. f_0 curves averaged across female and male speakers separately. Note the greater difference between sentences having the L tone and those having the H tone on their second syllables. Refer to panel R1C1 of Fig. 6 for the same curves averaged over all speakers.

whereas a low f_0 offset in the preceding tone (as in a L or F tone) lowers the maximum f_0 of the following tone.

When the other three variables are controlled for, variation in maximum f_0 due to the immediate carryover effect is as much as 28 Hz in syllable 3 (due to H and L tones in syllable 2), 24 Hz in syllable 4 (due to H and F tones in syllable 3), and 13 Hz in syllable 5 (due to H and L tones in syllable 4); variation in minimum f_0 due to the immediate carryover effect is as much as 20 Hz in syllable 3 (due to H and L tones in syllable 2), 10 Hz in syllable 4 (due to R and F tones in syllable 3), and 40 Hz in syllable 5 (due to H and L tones in syllable 4). Although smaller in magnitude than those due to tone proper and those due to focus, f_0 variations due to the immediate carryover effect are also quite extensive.

The pattern of long-distance carryover tonal influence (i.e. the influence of a tone on the f_0 of non-adjacent later syllables) seems to be more complicated than the immediate carryover influence. First, because every subject's recording was divided into sessions each having the same tone on syllable 2, the overall f_0 height varied from session to session. In particular, for unknown reasons, four of the speakers (three of them female) produced higher overall f_0 for sentences with H tone on syllable 2, and lower overall f_0 for sentences with L tone on syllable 2. One of the speakers did the opposite, while the other three produced roughly equal overall f_0 for both tonal conditions. As shown in Fig. 7, the average f_0 curves for male and female speakers differed in terms of variation of overall f_0 height. Female speakers' overall f_0 height was higher when the second syllable carried the H tone and lower when it carried the L tone. In contrast, male speakers' overall f_0 height varied much less.

A second complication about long-distance carryover effects can be seen in columns 13–14 of the top panel in Table VIa. The effects of the tone of syllable 2 on the maximum f_0 of syllable 4 was significant at the level of $p < 0.05$ ($F(3, 21) = 4.74$), when focus was neutral and when focus was on word 1. The direction of the effects, however, was rather

TABLE VII. Mean slopes of f_0 contour in syllable 1 averaged for 4 different tones on syllable 2 and for different focus conditions

	Tone of Syllable 2				Focus			
Subject	H	R	L	F	neutral	word 1	word 2	word 3
1	8.03	9.35	31.67	17.69	14.97	33.32	11.65	6.81
2	13.12	19.87	15.66	10.25	14.80	20.67	11.54	11.89
3	12.32	18.53	16.38	11.82	15.22	17.71	11.32	14.81
4	11.34	25.64	39.74	15.89	24.19	31.61	13.99	22.80
5	1.05	23.26	48.67	22.23	15.66	38.79	19.39	21.38
6	− 4.48	− 4.42	30.48	10.41	1.08	25.95	− 0.72	5.68
7	35.80	86.82	108.88	62.52	71.15	72.54	74.81	75.52
8	25.08	44.46	79.21	47.74	36.07	68.70	44.67	47.04
Mean	12.78	27.94	46.33	24.82	24.14	38.66	23.33	25.74

different in the two focus conditions. In the neutral focus condition, maximum f_0 of syllable 4 was higher when syllable 2 carried the H tone and lower when syllable 2 carried the L tone, a pattern consistent with the immediate carryover effects. When focus was on word 1, however, the highest maximum f_0 in syllable 4 occured when syllable 2 carried the L tone. This pattern is reminiscent of previous reports that the L tone, when being emphasized, raises the f_0 of the following tones (Shih, 1988; Shen, 1994; Xu, 1995).

Finally, in columns 17–20 of the top panel in Table VIa, the maximum f_0 of syllable 5 was influenced by the tone of syllable 2, but only in the neutral focus and final focus conditions. In columns 19–20 of the middle panel in Table VIa, the tone of syllable 3 influenced the maximum f_0 of syllable 5, when word 2 and word 3 were under focus. There thus seem to be some long-distance carryover influences, but their nature was mixed and further examination will be done later in the paper.

3.2.3. *Anticipatory effects*

As shown in Figs. 4–6, a tone does not affect the f_0 contour of the syllable before it as much as it does that of the syllable after it. The strongest anticipatory effect is seen in columns 9–12 of the bottom panel in both Table VIa and Table VIb, where the differences in maximum and minimum f_0 of syllable 3 due to the tone of syllable 4 (when the focus is on word 3) are as much as 16 Hz and 10 Hz, respectively. Also, the higher f_0 always occurs when syllable 4 carries the L tone. This agrees with the report of previous studies (Gandour *et al.*, 1992; Gandour *et al.*, 1994; Xu, 1993, 1994) that the anticipatory effect, when it occurs consistently, seems to be dissimilatory, i.e., a tone with a low pitch point raises rather than lowers the f_0 of the preceding tone.

In columns 1–4 of Table VIa and Table VIb, the maximum and minimum f_0 values of syllable 1 are shown to be barely affected by any immediate anticipatory influences. Further examination of the data from individual speakers, however, reveals that this again seems to be related to the variation between recording sessions. A closer look at the f_0 contours produced by individual speakers revealed that utterances with a L tone on syllable 2 had greater positive f_0 slopes on syllable 1 than those with a H tone on syllable 2, regardless of the overall f_0 height. The slope of f_0 contours in syllable 1 was therefore compared among utterances with different tones on syllable 2. The slopes were obtained by regressing all f_0 values in syllable 1 against time (f_0 was converted to

TABLE VIII. Differences in Hz between maximum f_0 of syllables 1 and 5 due to tone of syllable 2, 3 and 4 and focus conditions. The differences were computed by subtracting maximum f_0 of syllable 5 from that of syllable 1. The column headings indicate tones of the middle syllables in the sentence

	HHH	HHL	HRH	HRL	HFH	HFL	RHH	RHL	RRH	RRL	RFH	RFL
Neutral	0.9	26.4	7.8	29.2	7.9	27.6	7.7	28.0	13.7	32.1	9.9	30.0
Word 1	79.7	85.0	80.6	84.8	80.3	82.9	78.5	82.9	76.7	85.6	76.3	84.1
Word 2	45.1	51.2	43.5	57.0	52.5	52.6	49.2	53.4	39.7	53.1	49.9	54.1
Word 3	−12.8	12.7	−9.3	18.8	−10.4	13.1	−8.8	20.2	−3.7	18.8	−3.3	19.2

	LHH	LHL	LRH	LRL	LFH	LFL	FHH	FHL	FRH	FRL	FFH	FFL
Neutral	19.9	34.6	22.4	39.5	18.7	36.7	12.7	36.4	18.3	34.3	13.9	33.8
Word 1	80.4	89.1	76.4	89.9	82.7	90.2	80.0	83.5	80.0	83.1	78.7	85.5
Word 2	51.6	59.9	47.7	61.0	59.1	62.7	47.6	57.6	45.7	60.2	53.7	58.4
Word 3	1.9	20.9	6.1	26.9	2.2	25.0	−2.7	22.0	3.6	28.0	1.2	23.6

a logarithmic scale before calculating the slopes in order to normalize male and female differences). Mean slopes of all 8 speakers are listed in Table VII for four different tones on syllable 2 and for four focus conditions. As shown in Table VII, while most of the mean f_0 slopes are positive, the steepest ones occur before the L tone, and the shallowest ones occur before the H tone. Focus on word 1 also sharply increases the f_0 slope. A 2-factor repeated-measure ANOVA found the effects of both the tone of syllable 2 and focus highly significant ($F(3,21) = 11.88$, $p < 0.001$ and $F(3,21) = 11.63$, $p < 0.001$, respectively). The tone of the second syllable therefore seems to exert a rather strong anticipatory influence on the f_0 contour of the first syllable.

In columns 5–8 of the middle panel in Table VIa, significant anticipatory effects exerted by the tone of syllable 3 on the maximum f_0 of syllable 2 can be seen. The maximum and minimum f_0 values in syllable 2 are higher when followed by either the R or F tone and lower when followed by the H tone.

As for long-distance (non-contiguous) anticipatory effect, it can seen in Table VI that it is either absent or much smaller than that exerted by an adjacent tone. The only thing worth mentioning is the effect of syllable 4 on the maximum f_0 of syllable 1 ($F(1,7) = 9.46$, $p = 0.018$) (column 4, top panel of Table VIa), which, though not extensive, is consistent with the anticipatory influence of syllable 4 on other syllables (columns 5–12 of bottom panels in Tables VIa). Compared with the strong long-distance anticipatory raising reported for Yoruba (Laniran & Clements, 1995), the long-distance anticipatory raising seen here for Mandarin in quite small.

3.3. *Downtrends*

The analyses performed so far indicate that tone and focus can account for much of the variation in the shape and height of f_0 contours in short Mandarin declarative utterances. An additional question is how much of the overall downtrend, which is easily observable in Figs 3–6, can be attributed to these two factors. One indication of the overall downtrend is the absolute difference between the f_0 value at the beginning of the utterance and that at the end of the utterance. This difference can be measured as the difference between the peak f_0 values in the very first and very last syllables in an utterance. Such measurements are displayed in Table VIII. In the table, the differences in

Hz between the maximum f_0 values of syllables 1 and 5 are displayed according to the tones of syllables 2, 3, and 4 (in columns) and according to the focus conditions (in rows). The differences were computed by subtracting the maximum f_0 of syllable 5 from that of syllable 1. A 4-factor repeated-measure ANOVA found the effect on these difference values due to the four factors all significant. Tone of syllable 2: $F(3, 21) = 4.91$, $p < 0.01$; tone of syllable 3: $F(2, 14) = 4.97$, $p < 0.05$; tone of syllable 4: $F(1, 7) = 24.15$, $p < 0.01$; and focus: $F(3, 21) = 34.90$, $p < 0.01$. As shown in Table VIII, when there is no low tonal target in the sentence, as in the HHHHH case (column HHH in the table) and when there is no narrow focus in the sentence, the overall downtrend is less than 1 Hz (0.9). Non-H tones, which all have a low pitch region, all increase the difference between the maximum f_0 of syllable 1 and syllable 5. The amount of increase due to the L tone is much more than that due to the R and F tones, and the increases due to the R and F tones are roughly the same. Utterances with greater number of low pitch regions show greater differences than those with fewer low regions. In other words, the immediate carryover effect which was found to be highly significant, the long-distance carryover effects which were found earlier to be somewhat mixed, and the immediate anticipatory raising effect which was also found to be significant, are now seen as a combined effect which increases the f_0 difference between the very first and very last syllables in the utterance. In addition, focus on word 1 can increase the difference in sentences with a HHHHH pattern to as much as 79.7 Hz. Focus on word 3, in contrast, decreases the difference in HHHHH sentences by as much as -12.8 Hz.[3]

3.4. *Summary of focus and tonal effects*

To summarize the various tonal effects analyzed so far, the maximum f_0 of a tone is influenced by several factors: tone proper (as much as 43 Hz), focus (as much as 37 Hz), assimilatory carryover influence (as much as 28 Hz), and dissimilatory anticipatory influence (as much as 16 Hz). The minimum f_0 of a tone can vary due to tone proper (as much as 84 Hz), focus (as much as 26 Hz), assimilatory carryover influence (as much as 40 Hz), and dissimilatory anticipatory influence (as much as 10 Hz). Furthermore, when a tone occurs in a non-final word that is under focus, its f_0 contour expands extensively, and it exerts greater carryover and anticipatory influence on the adjacent tones and sometimes also on non-adjacent tones. When a tone occurs in a post-focus word, its f_0 contour is suppressed severely, and its influence on the surrounding tones is reduced.

3.5. *Tone-syllable alignment*

There are various ways an f_0 contour associated with a tone may conceivably be aligned with the syllable that carries the tone. It may be aligned with (a) the syllable onset, (b) the syllable offset, (c) the syllable center, or (d) the entire syllable. It may also be aligned with the syllable in some other ways, for example, with the P-center (Morton, Marcus &

[3] It could be the case that the f_0 difference shown in Table VIII is biased by the higher intrinsic pitch of the vowel /i/ in syllable 5. To check for this possibility, an ANOVA was conducted comparing the effect of vowel (/ao/ vs. /i/), focus (on-focus vs. off-focus), and position of the vowel in the sentence (early vs. later) on the maximum f_0 measured in the two syllables containing the diphthong /ao/ and the two syllables containing the vowel /i/. While the effects of both position ($F(1, 7) = 72.77$, $p < 0.001$) and focus ($F(1, 7) = 32.57$, $p < 0.001$) are highly significant, the effect of vowel is not ($F(1, 7) = 3.02$, $p = 0.126$), indicating that the intrinsic pitch difference was minimized in these sentences and did not significantly bias the f_0 measurements.

Frankish, 1976), as suggested by Hermes (1997). Or, f_0 contours and syllables may not be aligned with one another at all. If f_0 contours and syllables do align with one another, there should exist evidence for such alignment. First, there could be a region in the syllable where f_0 contours associated with a tone vary the least with the surrounding tones. Second, certain critical points in the f_0 contour might move synchronously with certain portions of the syllable, such as its onset or offset. The following analysis therefore looks for evidence for possible patterns of f_0-syllable alignment in terms of the region of least variability and alignment of critical f_0 points.

3.5.1. *Region of least variability*

To locate possible regions of least f_0 contour variability, f_0 curves can be displayed in such a way that f_0 variations due to surrounding tones can be visualized immediately. Such is already done in Figs. 4–6, as described earlier. Starting from Fig. 4, where in each panel only the second syllables carry different tones, it can be seen that f_0 contours of the first syllable exhibit little variation, regardless of the tone of the second syllable. In contrast, f_0 contours of the third syllable vary substantially with the tone of the second syllable. Closer inspection reveals that the early portion of the f_0 contour in the third syllable varies much more than the later portion. And, it looks as if the f_0 contours in the third syllable gradually converge while approaching the end of the syllable, although complete convergence is not quite achieved when the tone is H or R. Looking further at the direction of the f_0 variation in the third syllable, it becomes apparent that the f_0 height of the third syllable varies *with* that of the second syllable: the higher the latter, the higher the former.

Turning now to Fig. 5, where the *third* syllables in each panel carry different tones, the same tendencies can be seen as were found in Fig. 4. In this case, the tones of the first and second syllables both show little variation with the tone of the third syllable, whereas the tone of the fourth syllable exhibits much variation. Again, the f_0 variation is much greater in the early portion of the fourth syllable than in the later portion, and the gradual convergence of f_0 contours toward the end of the fourth syllable is evident in every panel in Fig. 5. Furthermore, the f_0 height of the fourth syllable varies *with* that of the second syllable.

The f_0 contour variation patterns in Fig. 6 do not appear, at a first glance, to be quite so similar to those in Fig. 4 and Fig. 5. When the fourth syllable has the L tone, the f_0 contour of the last syllable does not have the initial rise as seen in the H tone in earlier syllables when being preceded by a L tone. What makes the difference is probably the fact that the last syllable has an initial voiceless stop rather than a nasal when it is preceded by the L tone. It has been well established that an initial voiceless consonant raises the early portion of the f_0 contour in the syllable (Lehiste & Peterson, 1961; Hombert, 1978; Santen & Hirschberg, 1994). In most cases in Fig. 6, the initial voiceless stop [t] seems to sufficiently raise the early portion of the f_0 contour to offset part of the carryover influence of the preceding L tone, although in many cases the overall f_0 height in syllable 5 is still lowered by a preceding L tone.

In general, therefore, Figs. 4–6 demonstrate that the f_0 contour of a tone has the least variation toward the end of the syllable carrying the tone. This agrees with findings reported by Gandour *et al.* (1994) and Xu (1997) that there is greater carryover influence from a preceding tone than anticipatory influence from a following tone. Furthermore, assuming that high-level, rising, low, and falling are the contours appropriate for the correct perception of the four Mandarin tones as found by Whalen & Xu (1992), for each

tone, the most appropriate f_0 contour seems to occur in the later portion of the syllable that carries it. In panels C1RI-R2 and C4R1-R2 of Fig. 4, the contour of the H tone does not start to level off until near the end of the syllable. In panels C2R1-R2 and C5R1-R2 of Fig. 4, the f_0 contour of the R tone does not start to rise until near the middle of the vowel in the syllable. In Panels C3R1-R2 and C6R1-R2 of Fig. 4, the f_0 contour of the F tone starts to fall right after the initial consonant of the syllable when the preceding tone is H or R, but the fall begins much later when the preceding tone is L or F. And, in panels C4R1-C6R2 of Fig. 4, the f_0 contour of the L tone starts to fall from the onset of the initial consonant in the syllable, but the lowest point is not reached until the end of the syllable. It therefore seems that when two tones are produced in succession, the f_0 of the second syllable always starts from the ending f_0 of the first syllable, and proceeds from there toward an f_0 contour appropriate for the tone of the second syllable. This approximation process seems to continue until the end of the syllable. Consequently, as time elapses within the syllable, the f_0 contour varies less and less with the preceding tone and becomes more and more appropriate for the current tone.

3.5.2. *Alignment of syllable and f_0 contour*

The observations on the region of least variability seem to suggest that the implementation of a tone starts and ends with the syllable boundaries: starting at the syllable onset and ending at the syllable offset. The syllable boundaries thus appear to serve as the reference points for the alignment of tones. To determine to what extent this is the case, further alignment analysis is needed. In particular, the alignment between the syllable boundaries and certain critical points in the f_0 contours (such as peaks and valleys) needs to be examined. One way to do this is to plot the location of these points relative to the syllable onset as a function of syllable duration and examine the least-square regression line that best fits the function. A number of studies have done such analyses for f_0 peaks (e.g. Steele, 1986; Silverman & Pierrehumbert, 1990; Prieto *et al.*, 1995; Arvaniti & Ladd, 1995; Arvaniti *et al.*, 1998). In these studies, a measurement called "peak delay" is used. Peak delay measures the distance between an f_0 peak and the onset of the syllable or the rhyme associated with the peak. The peak delay is then further examined for its correlation with either the duration of the pitch-peak-bearing syllable (Prieto *et al.*, 1995; Kim, in press) or the duration of the vowel in the syllable (Steele, 1986; Silverman & Pierrehumbert, 1990). It has not been emphasized, however, that syllable duration is actually *equivalent* to the location of the sylllable offset relative to the syllable onset, and rhyme duration is *equivalent* to the location of syllable offset relative to the rhyme onset. Because of this, examining the location of the critical points as a function of syllable duration, in fact, reveals how these points align with syllable onset and offset, as illustrated in Fig. 8.

Shown in Fig. 8 are hypothetical regression plots that represent a number of possible alignment patterns between a critical f_0 point and a syllable. The dashed line in the figure has a slope of 1 and a y-intercept of 0. The points in cloud 4 are best fit by this line. This regression function indicates that these points move almost fully in synchrony with the syllable offset: no matter where the syllable offset is, the critical f_0 point always stays close to it. A regression function with a slope of 0.5 (as illustrated by cloud 2) indicates that these points maintain an equal distance from the syllable onset and offset. A regression function with a slope between 0.5 and 1 (as illustrated by cloud 3) indicates that the critical points move more in synchrony with syllable offset than with syllable onset,

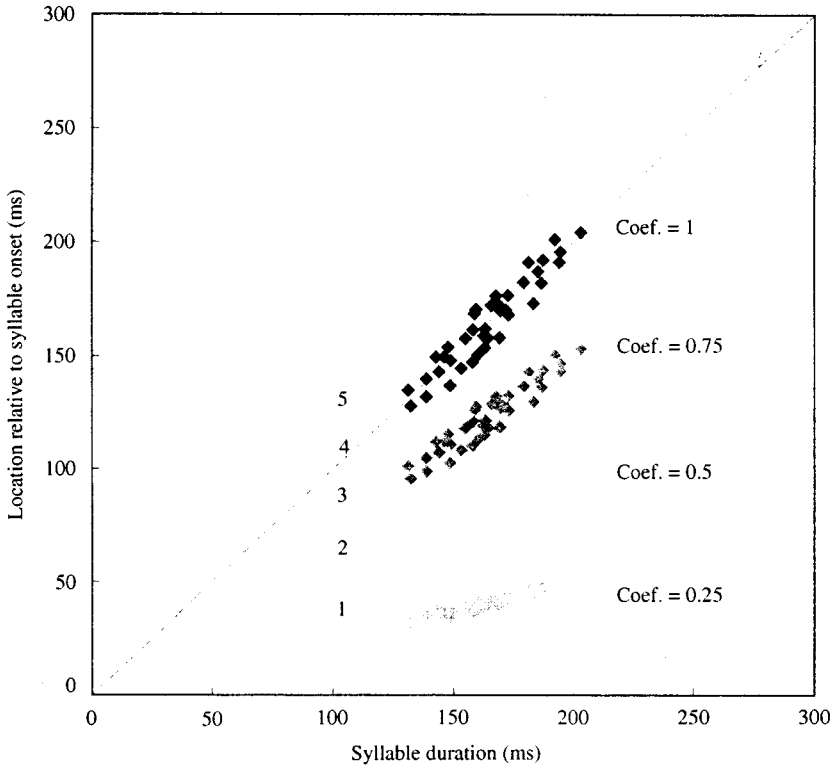

Figure 8. Hypothetical relations between f_0-peaks and syllable onset and offset, revealed by plotting f_0-peak location as a function of syllable duration: Cloud 4: Coefficient = 1; f_0 peaks move with syllable offset when syllable duration changes; Cloud 2: Coefficient = 0.5; f_0 peaks maintain an equal distance from the syllable onset and offset; Cloud 3: 0.5 < Coefficient < 1; f_0 peaks move more in synchrony with syllable offset than with syllable onset; Cloud 1: 0 < Coefficient < 0.5; f_0 peaks move more in synchrony with syllable onset than with syllable offset; Cloud 5: f_0 peaks associated with a particular syllable are delayed slightly beyond the syllable offset.

whereas a regression function with a slope between 0 and 0.5 (as illustrated by cloud 1) indicates that these critical points move more in synchrony with syllable onset than with syllable offset. In some cases, the critical points associated with a syllable are delayed beyond the syllable offset. When that happens, the distribution of these points as a function of syllable duration may look like cloud 5 in Fig. 8. These distribution patterns are used as references in the regression analyses to be discussed below, together with regression analyses involving non-positional measurements such as f_0 height and slope.

Alignment of syllable and f_0 contour — right edge. To be examined first is the alignment of the right edge of the f_0 contours in the R tone. When this tone is followed by a tone with a low onset (L or R), the rising movement seems to continue past the offset of the R-tone-carrying syllable, and the contour does not start to fall until the middle of the initial consonant of the following syllable, as is seen in panels C5R1-R2 of Fig. 4, panels C1R5-C4R6 of Fig. 5, and panels C1R5-C4R6 of Fig. 6. To examine this alignment more

TABLE IX. Mean and standard deviation at measurements taken from the R tone in word 2 that is followed by a L tone

	Duration	Slope	Maxf$_0$	Onset-to-maxf$_0$	Offset-to-maxf$_0$
M	198 ms	108	201 Hz	228 ms	29 ms
SD	24.7	26.6	77.4	26.0	13.6

Note: In this and the following tables, the slopes were obtained by regressing f_0 of all data points in syllable 1 against time. The f_0 values were converted to a logarithmic scale before calculating the slopes in order to normalize male and female f_0 differences.

closely, several measurements were taken around the third syllable in the utterances: (1) duration of the R-tone-carrying syllable (duration), (2) slope of the rising contour from its lowest point to syllable offset (slope), (3) maximum f_0 in the rising contour (maxf$_0$), (4) location of maximum f_0 relative to the onset of the R-tone-carrying syllable (onset-to-maxf$_0$), and (5) location of maximum f_0 relative to the offset of the R-tone-carrying syllable (offset-to-maxf$_0$). In some utterances a turning point could not be observed near the end of the R-tone-carrying syllable. This was the case when the fourth syllable had a H tone, and when focus was on the first word (to be discussed later). That left 60 utterances from which these measurements were taken. Table IX displays the means and standard deviations of these measurements across the eight subjects. Univariate linear regression analyses were performed on these measurements for each subject. Table X lists the individual regression coefficients, mean coefficients, estimated standard deviations (SE), results of two-tailed single group *t* tests, and mean correlation coefficients (*r*).

As shown in Table IX, the mean offset-to-maxf$_0$ is 29 ms. This indicates that many f_0 peaks in the R tone probably indeed occurred after the syllable offset. In Table X, the mean correlation between duration and onset-to-maxf$_0$ is 0.91—much higher than other mean *r* values. Such a high correlation indicates that the f_0 peak probably moves closely *with* the syllable offset: the farther away the syllable offset is from the syllable onset (which is equivalent to greater duration), the farther away the location of the f_0 peak is from the syllable onset. This is verified by the regression analyses with duration as the predictor and onset-to-maxf$_0$ as the dependent variable. The regression coefficients are all very close to 1, as can be seen in the table. In contrast to the high correlation between duration and onset-to-maxf$_0$, offset-to-maxf$_0$ is not strongly correlated with anything except onset-to-maxf$_0$ (which is equivalent to duration + offset-to-maxf$_0$). This indicates that the distance between syllable offset and f_0 peak did not vary with anything else systematically. In other words, f_0 peak in the R tone simply stayed close to the syllable offset.

Table X also shows, however, that when onset-to-maxf$_0$ is regressed on duration, most of the coefficients are slightly greater than 1. This indicates that the distance between syllable onset and f_0 peak (onset-to-maxf$_0$) actually increased faster than the increase of syllable duration. In other words, the f_0 peak occurred increasingly later than the syllable offset as syllable duration increased (thus differing slightly from cloud 5 in Fig. 8, where the regression coefficient is 1). As shown in Table X, onset-to-maxf$_0$ is also correlated with maximum f_0 (maxf$_0$) and slope of the rising contour (slope). Since both maxf$_0$ and slope correlate with duration, it is likely that the faster increase of onset-to-maxf$_0$ than the increase of duration is related to the increased maxf$_0$ as well as slope. This

TABLE X. Univariate linear regression analyses of key variables shown in Table IX for each of the eight subjects. The measurements were taken from the R tone in word 2 which is followed by a L tone

	Duration				Slope			$Maxf_0$		Onset-to-$maxf_0$
Subject	Slope	$Maxf_0$	Onset-to-$maxf_0$	Offset-to-$maxf_0$	$Maxf_0$	Onset-to-$maxf_0$	Offset-to-$maxf_0$	Onset-to-$maxf_0$	Offset-to-$maxf_0$	Offset-to-$maxf_0$
1	0.46[b]	0.26[b]	1.09[b]	0.09[a]	0.27[b]	0.55[b]	0.01	1.28[b]	0.06	0.13[b]
2	0.53[b]	0.25[b]	1.26[b]	0.26[a]	0.20[b]	0.60[b]	0.22[a]	1.20[b]	0.23	0.38[b]
3	0.68[b]	0.37[a]	1.05[b]	0.05	0.58[b]	0.45[b]	0.20[b]	0.43[b]	0.22[b]	0.32[b]
4	0.26	0.18[b]	1.18[b]	0.18[b]	− 0.05	0.31[a]	0.09	1.46[b]	0.19	0.24[b]
5	1.90[b]	0.42[b]	1.18[b]	0.18	0.22[b]	0.54[b]	0.13[b]	1.97[b]	0.49[b]	0.38[b]
6	0.72[b]	0.20[b]	1.09[b]	0.09[a]	0.17[b]	1.00[b]	0.08	2.32[b]	0.33[a]	0.16[b]
7	1.77[b]	0.29[b]	1.18[b]	0.18[a]	0.09[b]	0.38[b]	0.09[b]	2.03[b]	0.42[a]	0.32[b]
8	0.77[b]	0.29[b]	0.96[b]	− 0.04	0.17[b]	0.29[b]	− 0.03	1.43[b]	− 0.09	0.05
Mean	0.89	0.28	1.12	0.12	0.21	0.52	0.10	1.51	0.23	0.25
SE	0.22	0.03	0.03	0.03	0.06	0.08	0.03	0.21	0.07	0.04
$t(7)$	4.11	9.88	33.33	3.69	3.31	6.40	3.34	7.21	3.47	5.64
p	0.005	0.000	0.000	0.008	0.013	0.000	0.012	0.000	0.010	0.001
mean r	0.56	0.57	0.91	0.23	0.52	0.58	0.24	0.57	0.21	0.58

Note: Variables in the column spanners are treated as independent variables and those in column headers as dependent variables. t and p are results of two-tailed single-group t tests against a mean of 0 (see Lorch and Myers, 1990).

The last row shows the mean correlation cofficient averaged over subjects.

[a] $p < 0.05$ for the simple regression model; [b] $p < 0.01$.

TABLE XI. Mean and standard deviation of measurements taken from the F tone in word 2 followed by a H tone. Sentences with focus on word 2 are excluded

	Duration	Slope	$Minf_0$	Onset-to-$minf_0$	Offset-to-$minf_0$
M	184 ms	− 87	163 Hz	191 ms	7 ms
SD	22. 8	23.3	65.3	15.5	11.5

interpretation was confirmed by separate correlation analyses on utterances whose focus was on word 2 (i.e., syllable 3) and utterances whose focus was not on word 2. While in both cases the correlation between duration and onset-to-$maxf_0$ remained high to moderate (mean $r = 0.82$, $t(7) = 2.90$, $p < 0.001$, when focus was on word 2; mean $r = 0.57$, $t(7) = 2.85$, $p < 0.001$, when focus was not on word 2), the slopes of the regression line were reduced to less than 1.0 in both cases (0.821, when focus was on word 2; 0.823, when focus was not on word 2). Furthermore, the means of offset-to-$maxf_0$ and $maxf_0$ were 24 ms and 194 Hz, when word 2 was not under focus, but 38 ms and 213 Hz, when it was under focus. Apparently, a focus both increased the maximum f_0 in the R tone and pushed its f_0 peak further away from the syllable offset.

Similar to the R tone, the F tone also has a rapid f_0 movement in the final portion of its contour. To see if the alignment of the right edge in the F tone is similar to that of the R tone, comparable measurements were taken from the F tone in word 2 (syllable 3) when followed by a H tone: (1) duration of the F-tone-carrying syllable (duration), (2) slope of the falling contour from its highest point to syllable offset (slope), (3) minimum f_0 in the falling contour ($minf_0$), (4) location of minimum f_0 relative to the onset of the F-tone-carrying syllable (onset-to-$minf_0$), and (5) location of minimum f_0 relative to the offset of the F-tone-carrying syllable (offset-to-$minf_0$). (Because having a focus on the second word lowered the f_0 of the following tone extensively, it was not always possible to locate an f_0 valley, and as a result only 40 utterances could be used.) Table XI displays the means and standard deviations of these measurements across eight subjects. As can be seen in Table XI, the mean offset-to-$minf_0$ is 7 ms, indicating that the f_0 valley in the F tone occurred in general *after*, but very close to the syllable offset. To examine the alignment more carefully, simple regression analyses between these measurements were performed for each speaker. Table XII displays regression coefficients and mean correlation coefficients obtained in these analyses. As shown in Table XII, again, offset-to-$minf_0$ is not correlated with anything except onset-to-$minf_0$ (which is equivalent to duration + offset-to-$minf_0$). In contrast, onset-to-$minf_0$ is significantly correlated with duration (mean $r = 0.66$). At the same time, the mean regression coefficient for duration on onset-to-$minf_0$ is close to 1 (0.90, thus similar to cloud 5 in Fig. 8). Therefore, similar to the f_0 peak in the R tone, the f_0 valley in the F tone not only occurred in general after the syllable offset, but also stayed close to and moved in synchrony with the syllable offset.

Alignment of syllable and f_0 contour — left edge. The finding that the location of the final extreme f_0 in a tone stayed close to syllable offset does not say anything about how the early critical point of a tone is aligned with the syllable. To examine this alignment, another set of measurements was taken for the R and F tones in word 2. They are (1) duration of the target-tone-carrying syllable, (2) slope of the f_0 contour from the

TABLE XII. Univariate linear regression analyses of key variables shown in Table XI for each of the eight subjects

	Duration				Slope			$Minf_0$		Onset-to-$minf_0$
Subject	Slope	$Minf_0$	Onset-to-$minf_0$	Offset-to-$minf_0$	$Minf_0$	Onset-to-$minf_0$	Offset-to-$minf_0$	Onset-to-$minf_0$	Offset-to-$minf_0$	Offset-to-$minf_0$
1	− 0.79	0.05	0.82[b]	− 0.18	− 0.8[a]	− 2.22[d]	− 0.13[a]	0.59	0.47	0.61[b]
2	0.13	0.11	0.83[b]	− 0.17	− 0.30[b]	− 0.02	− 0.09	0.50	− 0.49[a]	− 0.67[b]
3	0.25	− 0.01	1.21[b]	0.21	0.52[b]	− 0.06	− 0.08	− 0.11	− 0.11	0.60[b]
4	0.54	− 0.65[b]	0.63[b]	− 0.37[a]	0.04	0.13	− 0.02	− 0.23	0.28	0.46[b]
5	− 0.75[a]	− 0.03	0.98[b]	− 0.02	− 0.04	− 0.06	0.08	− 0.25	− 0.01	0.33[b]
6	− 0.40	− 0.08	0.89[b]	− 0.11	− 0.07[a]	− 0.03	0.08	− 0.78	− 0.30	0.18[a]
7	− 0.68	− 0.20[a]	1.09[b]	0.09	0.12[b]	− 0.11	− 0.04	− 0.47	0.11	0.49[b]
8	3.14[b]	0.34[b]	0.71[b]	− 0.29	0.08[b]	0.10[a]	− 0.03	1.39[b]	0.02	0.68[b]
Mean	0.18	− 0.07	0.90	− 0.10	0.03	− 0.03	− 0.03	0.08	0.12	0.50
SE	0.46	0.10	0.07	0.07	0.08	0.04	0.03	0.25	0.10	0.06
$t(7)$	0.40	0.73	13.22	1.55	0.41	0.86	1.09	0.33	1.21	8.05
p	0.705	0.492	0.000	0.166	0.695	0.417	0.311	0.753	0.264	0.000
Mean r	0.01	− 0.05	0.66	− 0.11	0.09	− 0.05	− 0.04	0.01	0.07	0.66

Note: Variables in the column spanners are treated as independent variables and those in column headers as dependent variables. t and p are results of two-tailed single-group t tests against a mean of 0 (see Lorch and Myers, 1990).

The last row shows the mean correlation coefficient averaged over subjects.

[a] $p < 0.05$ for the simple regression model; [b] $p < 0.01$.

TABLE XIII. Mean and standard deviation of measurements taken from the R tone in word 2 preceded by a H or R tone

	Duration	Slope	$Minf_0$	Onset-to-$minf_0$	$Minf_0$-to-offset
M	198 ms	85	158 Hz	117 ms	81 ms
SD	23.3	23.1	59.1	11.2	13.9

TABLE XIV. Mean and standard deviation of measurements taken from the F tone in word 2 preceded by a H tone

	Duration	Slope	$Maxf_0$	Onset-to-$maxf_0$	$Maxf_0$-to-offset
M	200 ms	− 107	204 Hz	117 ms	83 ms
SD	19.3	29.1	77.0	10.7	19.8

lowest or highest point to syllable offset, (3) minimum and maximum f_0 in the f_0 contour ($minf_0$ and $maxf_0$), (4) location of $maxf_0$ and $minf_0$ relative to syllable onset (onset-to-$maxf_0$ and onset-to-$minf_0$), and (5) location of $maxf_0$ and $minf_0$ relative to syllable offset (max-to-offset and min-to-offset). The means and standard deviations of these measurements are displayed in Tables XIII and XIV for the R and F tones, respectively.

In Table XIII, the mean onset-to-$minf_0$ is 117 ms, while the mean $minf_0$-to-offset is 81 ms. This indicates that the location of the f_0 valley in the R tone is, on average, closer to the syllable offset than to the onset. Similarly, in Table XIV the mean onset-to-$maxf_0$ is 117 ms, while the mean $maxf_0$-to-offset is 83 ms, indicating that the location of the f_0 peak in the F tone is on average closer to the syllable offset than to the onset. To more closely examine the alignment of the early critical points in the R and F tones, simple regression analyses on these measurements were performed for each speaker. Tables XV and XVI display regression coefficients and mean correlation coefficients obtained in these analyses for the R and F tones respectively. In Table XV, both onset-to-$minf_0$ and $minf_0$-to-offset are significantly correlated with syllable duration ($r = 0.76$ and 0.77). This means that as the syllable duration increased, the distances between minimum f_0 and syllable onset and between minimum f_0 and syllable offset both increased, indicating that the location of the f_0 valley stayed near the middle of the syllable, though somewhat closer to syllable offset than to the onset, thus corroborating the means in Table XIII. Furthermore, the mean regression coefficient with duration as the predictor is 0.50 when onset-to-$minf_0$ is the dependent variable. This resembles the regression coefficient for cloud 2 in Fig. 8, indicating that the location of the f_0 valley probably remained near the center of the syllable as the syllable duration increased. This is further verified by the mean regression coefficients when $minf_0$-to-offset is the dependent variable (0.50), indicating that the distances between the f_0 valley and syllable onset and between the f_0 valley and syllable offset both increased half as fast as syllable duration.

In Table XVI, while onset-to-$maxf_0$ is significantly correlated with duration, $maxf_0$-to-offset is not, indicating that for the F tone, the initial f_0 peak moved more in synchrony with syllable offset than with syllable onset. This is supported by the regression coefficients with duration as the predictor: 0.65 when onset-to-$maxf_0$ is the dependent variable, and 0.35 when $maxf_0$-to-offset is the dependent variable. The fact that the f_0 peak is

TABLE XV. Univariate linear regression analyses of key variables shown in Table XIII for each of the eight subjects

	Duration				Slope			$Minf_0$		Onset-to-$minf_0$
Subject	Slope	$Minf_0$	Onset-to-$minf_0$	$Minf_0$-to-offset	$Minf_0$	Onset-to-$minf_0$	$Minf_0$-to-offset	Onset-to-$minf_0$	$Minf_0$-to-offset	Offset-to-$minf_0$
1	-0.82^b	0.25^b	0.68^b	0.32^b	-0.07	0.44^b	0.12^a	-0.85^b	-0.43^b	0.15
2	1.45^b	-0.44^b	0.67^b	0.33^b	-0.20^b	0.32^b	0.07	-0.90^b	-0.26^a	-0.05
3	0.99^b	-0.25^b	0.54^b	0.46^b	-0.11^b	0.20^b	0.14^b	-0.44^a	-0.99^b	0.05
4	0.42^b	-0.46^b	0.53^b	0.47^b	-0.29^b	0.38^b	0.07	-0.82^b	-0.67^b	0.32^a
5	1.89^b	-0.12^b	0.51^b	0.49^b	-0.04^b	0.20^b	0.21^b	-1.62^b	-1.01^b	0.48^b
6	1.00^b	-0.13^b	0.34^b	0.66^b	-0.12^b	0.21^b	0.56^b	-1.35^b	-3.28^b	0.57^b
7	1.84^b	-0.38^b	0.48^b	0.52^b	-0.15^b	0.14^b	0.20^b	-0.94^b	-0.76^b	0.19
8	1.50^b	-0.19^b	0.24^b	0.76^b	-0.10^b	0.04	0.31^b	-0.28	-1.01^b	-0.04
Mean	1.24	-0.28	0.50	0.50	-0.13	0.24	0.21	-0.90	-1.05	0.21
SE	0.18	0.05	0.05	0.05	0.03	0.05	0.06	0.15	0.33	0.08
$t(7)$	6.78	5.81	9.31	9.42	4.79	5.16	3.66	5.84	3.14	2.57
p	0.000	0.001	0.000	0.000	0.002	0.001	0.008	0.001	0.016	0.037
Mean r	0.711	-0.67	0.76	0.77	-0.53	0.55	0.51	-0.52	-0.50	0.20

Note: Variables in the column spanners are treated as independent variables and those in column headers as dependent variables. t and p are results of two-tailed single-group t tests against a mean of 0 (see Lorch and Myers, 1990).

The last row shows the mean correlation coefficient averaged over subjects.

[a] $p < 0.05$ for the simple regression model; [b] $p < 0.01$.

TABLE XVI. Univariate linear regression analyses of key variables shown in Table XIV for each of the eight subjects

	Duration				Slope			$Maxf_0$		Onset-to-$minf_0$
Subject	Slope	$Maxf_0$	Onset-to-$maxf_0$	$Maxf_0$-to-offset	$Maxf_0$	Onset-to-$maxf_0$	$Maxf_0$-to-offset	Onset-to-$maxf_0$	$Maxf_0$-to-offset	Offset-to-$maxf_0$
1	-1.19^b	0.80^b	0.41^b	0.59^b	-0.56^b	-0.25^b	-0.14^a	0.29^b	0.26^b	-0.23
2	-0.97^b	0.63^b	0.55^b	0.45^b	-0.51^b	-0.48^b	-0.10	0.22	0.58^b	-0.47^b
3	-1.24^b	0.74^b	0.52^b	0.48^b	-0.30^b	-0.14^b	-0.11^a	0.58^b	0.05	-0.37^b
4	-0.57^b	0.35^b	0.65^b	0.35^b	-0.58^b	-0.59^b	-0.29^b	0.29	0.48^b	0.10
5	-1.78^b	0.62^b	0.73^b	0.27^b	-0.31^b	-0.28^b	-0.02	0.87^b	0.09	-0.19^a
6	-1.83^b	0.42^b	0.68^b	0.32^b	-0.21^b	-0.32^b	-0.01	1.02^b	0.26	-0.34^b
7	-2.35^b	0.48^b	1.04^b	-0.04	-0.18^b	-0.33^b	-0.12^b	0.98^b	-0.25	-0.54^b
8	-1.94^b	0.65^b	0.66^b	0.34^b	-0.25^b	-0.13^b	-0.08^a	0.50^b	0.29^a	-0.40^b
Mean	-1.49	0.59	0.65	0.35	-0.36	-0.32	-0.08	0.59	0.22	-0.31
SE	0.21	0.06	0.07	0.07	0.06	0.06	0.04	0.11	0.09	0.07
$t(7)$	7.17	10.62	9.89	5.25	6.29	5.69	1.91	5.18	2.36	4.32
p	0.000	0.000	0.000	0.001	0.000	0.001	0.097	0.001	0.050	0.003
Mean r	0.56	0.57	0.91	0.23	0.52	0.58	0.24	0.57	0.21	0.58

Note: Variables in the column spanners are treated as independent variables and those in column headers as dependent variables. t and p are results of two-tailed single-group t tests against a mean of 0 (see Lorch and Myers, 1990).

The last row shows the mean correlation coefficient averaged over subjects.

[a] $p < 0.05$ for the simple regression model; [b] $p < 0.01$.

TABLE XVII. Mean and standard deviation of measurements taken from the H tone in word 2 that is preceded by a L or F tone and followed by a L tone. Sentences with focus on word 2 are excluded

	Duration	$Maxf_0$	Onset-to-$maxf_0$	$Maxf_0$-to-offset
M	199 ms	213 Hz	175 ms	24 ms
SD	29.4	79.9	25.4	15.9

TABLE XVIII. Mean and standard deviation of measurements taken from the L tone in word 1 that is followed by a H or F tone

	Duration	$Minf_0$	Onset-to-$minf_0$	$Minf_0$-to-offset
M	158 ms	123 Hz	155 ms	2 ms
SD	17.5	46.9	18.6	10.7

farther away from the syllable onset than from the offset (117 ms vs. 83 ms, as mentioned earlier) further confirms the closer proximity of the f_0 peak to syllable offset than to syllable onset.

Alignment of syllable and f_0 contour —H and L tones. As can be seen in Figs. 4–6, when a H tone is preceded by a F or L tone and followed by a L tone, the f_0 contour does not usually show a high plateau in the H-tone-carrying syllable. Instead, there is usually a f_0 peak near the end of the syllable. Similarly, there is usually a f_0 valley near the end of a L-tone-carrying syllable when it is before a H tone. To assess how the f_0 peak aligns with the H-tone-carrying syllable and how the f_0 valley aligns with the L-tone-carrying syllable, the following measurements were taken: (1) duration of the target-tone-carrying syllable, (2) maximum or minimum f_0 in the f_0 contour ($maxf_0$ or $minf_0$), (3) location of $maxf_0$ or $minf_0$ relative to syllable onset (onset-to-$maxf_0$ or onset-to-$minf_0$), and (4) location of $maxf_0$ or $minf_0$ relative to syllable offset ($maxf_0$-to-offset or $minf_0$-to-offset). Means and standard deviations of these measurements are shown in Tables XVII and XVIII for the H and L tones, respectively. As can be seen in the two tables, unlike the R tone (or the F tone) whose f_0 peak (or valley) in general occurred *after* the syllable offset, f_0 peaks or valleys in the H or L tone generally occurred *before* syllable offset, but still far way from syllable onset (24 vs. 175 ms in the H tone; and 2 vs. 155 ms in the L tone).

To more closely examine the alignment of the f_0 peak and valley in the H and L tones, simple regression analyses between the measurements were performed for each speaker. Tables XIX and XX display regression coefficients and mean correlation coefficients obtained in these analyses for the H and L tones, respectively. In Table XIX, onset-to-$maxf_0$ is strongly correlated with duration ($r = 0.76$), whereas $maxf_0$-to-offset is more weakly correlated with duration ($r = 0.38$). Using syllable duration as the predictor, the regression coefficient is 0.74 (similar to cloud 3 in Fig. 8) when onset-to-$maxf_0$ is the dependent variable, but 0.26 when onset-to-$maxf_0$ is the dependent variable. This indicates that the f_0 peak in the H tone moved more in synchrony with syllable offset than with the onset. As can be seen in Table XX, the alignment of the f_0 valley in the

TABLE XIX. Univariate linear regression analyses of key variables shown in Table XVII for each of the eight subjects

	Duration			$Maxf_0$		Onset-to-$maxf_0$
Subject	$Maxf_0$	Onset-to-$maxf_0$	$Maxf_0$-to-offset	Onset-to-$maxf_0$	$Maxf_0$-to-offset	$Maxf_0$-to-offset
1	0.24[b]	0.85[b]	0.15[b]	0.69	0.42[b]	0.07
2	0.65[b]	0.79[b]	0.21	0.82[b]	0.18	−0.37[b]
3	0.29[a]	0.56[a]	0.44[a]	0.88[b]	−0.39	−0.64[b]
4	0.26[b]	0.60[b]	0.40[b]	0.89[b]	0.02	0.10
5	0.61[b]	0.78[b]	0.22[a]	0.77[b]	0.03	−0.11
6	0.37[b]	0.65[b]	0.35[a]	1.53[b]	0.38	−0.26
7	0.48[b]	0.98[b]	0.02	1.29[b]	−0.01	−0.13
8	0.38[b]	0.72[b]	0.28[a]	1.27[b]	0.33	−0.26
Mean	0.41	0.74	0.26	1.02	0.12	−0.20
SE	0.06	0.05	0.05	0.11	0.09	0.09
(t) 7	7.43	15.14	5.25	9.51	1.27	2.38
p	0.000	0.000	0.001	0.000	0.243	0.049
mean (r)	0.66	0.76	0.38	0.62	0.13	−0.26

Note: Variables in the column spanners are treated as independent variables and those in column headers as dependent variables. t and p are results of two-tailed single-group t tests against a mean of 0 (see Lorch and Myers, 1990).

The last row shows the mean correlation coefficient averaged over subjects.

[a] $p < 0.05$ for the simple regression model; [b] $p < 0.01$.

TABLE XX. Univariate linear regression analyses of key variables shown in Table XVIII for each of the eight subjects

	Duration			$Minf_0$		Onset-to-$minf_0$
Subject	$Minf_0$	Onset-to-$minf_0$	$Minf_0$-to-offset	Onset-to-$minf_0$	$Minf_0$-to-offset	$Minf_0$-to-offset
1	-1.34^b	0.82^b	0.18	-0.25^b	0.10^b	-0.44^b
2	-0.91^b	0.63^b	0.37^a	-0.24^b	0.14^b	-0.70^b
3	−0.25	0.72^b	0.28^a	−0.10	−0.07	-0.65^b
4	-0.94^b	1.10^b	−0.10	-0.59^b	0.20^b	-0.36^b
5	-0.17^b	0.86^b	0.14	-1.26^b	0.27	-0.64^b
6	-0.23^b	0.51^b	0.49^b	-0.86^b	0.02	-0.64^b
7	-0.30^b	0.80^b	0.20^a	-1.32^b	0.24	-0.29^b
8	0.01	0.12	0.88^b	−0.37	0.43	-0.88^b
Mean	−0.52	0.69	0.31	−0.62	0.17	−0.58
SE	0.17	0.10	0.10	0.17	0.05	0.07
$(t)7$	3.05	6.80	3.00	3.73	3.03	8.38
p	0.019	0.000	0.020	0.007	0.019	0.000
mean (r)	−0.37	0.54	0.25	−0.43	0.17	−0.65

Note: Variables in the column spanners are treated as independent variables and those in column headers as dependent variables. t and p are results of two-tailed single-group t tests against a mean of 0 (see Lorch and Myers, 1990).

The last row shows the mean correlation coefficient averaged over subjects.

[a] $p < 0.05$ for the simple regression model; [b] $p < 0.01$.

L tone is very similar to that of the f_0 peak in the H tone. That is, the f_0 valley occurred before but very close to the syllable offset, and it moved more in synchrony with syllable offset than with the onset (mean regression coefficient: 0.69 vs. 0.31).

3.5.3. *Summary of tone syllable alignment*

To summarize the various alignment analyses, the results reveal the following alignment patterns in the short Mandarin sentences examined in this study: (a) The final portion (rather than other regions) of the f_0 contour of a syllable varies the least under the carry-over tonal influence; (b) the location of the later extreme f_0 point associated with a tone stayed close to the syllable offset; (c) the location of the earlier extreme f_0 point associated with a tone stayed roughly in the middle of the syllable, but in general closer to the syllable offset than to the onset; (d) as the syllable duration changed, the latter extreme point of a tone tended to move in synchrony with syllable offset, rather than with the onset: and (e) in general, the earlier critical point of a tone moved more in synchrony with the syllable offset than with the onset.

4. General discussion

Three major questions about the formation of f_0 contours in Mandarin were raised at the outset of this study. First, how tone and focus can be implemented simultaneously when both have to use fundamental frequency as a major acoustic correlate; second, how lexical tones can be adequately implemented acoustically when they have to interfere with one another as has been widely reported; and third, how the interaction of tone and focus and the interaction among the tones themselves affect the shape of f_0 contours and their alignment with the syllabic elements of an utterance. The following discussion will address these questions based on the results of the data analyses described in the previous sections.

To guide the discussion, selected mean f_0 curves produced by the male speakers are displayed in Fig. 9 to illustrate the main points (which can be corroborated by the f_0 tracings in Figs. 3–6). Displayed in the figure are the mean f_0 tracings of the sentences "māomī mō māomī" and "māomǐ mō mǎdāo" averaged across the four males speakers. The two sentences have the tone sequences of HHHHH and HLHLH, respectively. The four panels in Fig. 9 display the f_0 tracings of these sentences produced under four focus conditions: neutral focus (upper left), focus on word 1 (lower left), focus on word 2 (lower right), and focus on word 3 (upper right).

4.1. *How tones and focus are implemented in parallel*

The f_0 tracings in Fig. 9 illustrate the major effects of both focus and tone when other factors are kept constant. In general, tone identities are implemented as local f_0 contours, while focus patterns are implemented as pitch range variations imposed on different regions of an utterance. The pitch range of tonal contours directly under focus is substantially expanded; the pitch range after the focus is severely suppressed (lowered and compressed); and the pitch range before the focus does not deviate much from the neutral-focus condition. Thus, there seem to be three distinct focus-related pitch ranges: *expanded* in non-final focused words, *suppressed* (lowered and compressed) in post-focus

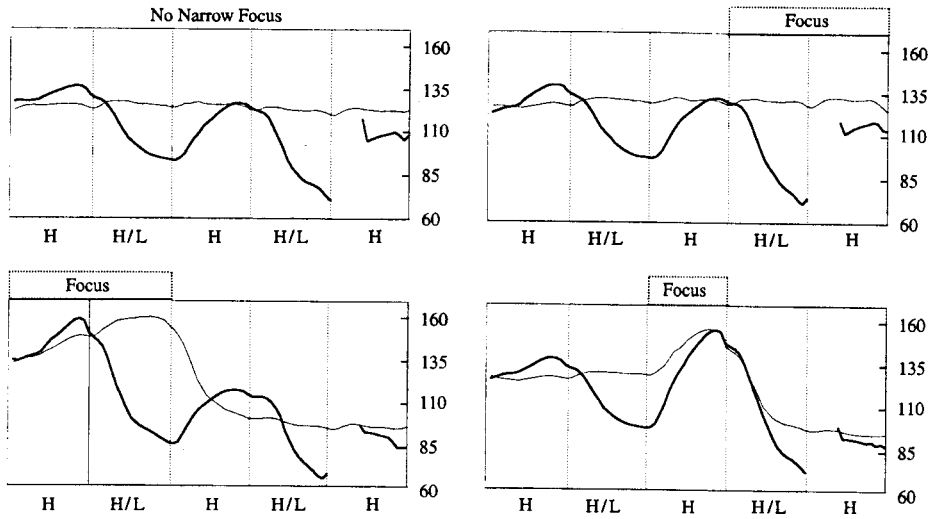

Figure 9. Selected mean f_0 curves of two sentences produced by the male speakers. The sentences have the tone sequences of HHHHH (thin line) and HLHLH (thick line), respectively. The f_0 curves in each graph were produced under one of the four focus conditions: neutral focus (upper left), focus on word 1 (lower left), focus on word 2 (lower right), and focus on word 3 (upper right).

words, and *neutral* in all other words. Fig. 9 also demonstrates that the effect of focus is much more than just fine adjustment of the local tone contours. Rather, the adjustments are fairly substantial. In fact, as discussed earlier, the effect of focus on the f_0 of a syllable is second only to the effect of the lexical tone carried by that syllable, and it is much greater than both anticipatory and carryover tonal influences. In the post-focus region, the pitch range is sometimes suppressed so severely that different tone contours are hardly distinct from one other.

An interesting aspect of the focus patterns is that, at least in production, the full realization of a focus seems to require that the f_0 of all words after the focus be suppressed. If, however, there is nothing to suppress, as is the case when focus is on the last word, on-focus expansion cannot be implemented effectively, thus leaving these supposedly narrow-focused utterances not much different from those with no narrow focus, as shown in Fig. 9, and as reported by Cooper *et al.* (1985) and Jin (1996). Previous *perception* tests have also found that a final narrow focus does not sound very different from a broad focus (Jin, 1996). This may explain the observation that in many languages, a sentence with no narrow focus is often described as having a final stress or a final nuclear tone. Due to the lack of reliable perceptual cues in f_0 that can separate a neutral (or broad) focus from a final focus (Bunnel, Hoskins & Yarrington, 1997), it is understandable that such confusion should occur.

An important implication of the findings about the effects of focus is that it is important for tone and intonation studies to keep focus under control, because without deliberate control one may run the risk of letting the subjects freely produce focus patterns as they see fit, thus inadvertently introducing substantial random effects into the f_0 contours under study. Another implication is that the asymmetry about a focus may extensively tilt the f_0 curve downward over the entire utterance when the focus occurs

before the last word. As demonstrated by Eady and Cooper (1986), in English, a large portion of the observed f_0 decline can be accounted for by the f_0 drop within a focused word that occurs early in a sentence. It is thus crucial that any study that looks at various downtrends in intonation also control for the effect of focus.

4.2. *How lexical tones maintain their identity while interacting with one another*

The f_0 tracings shown in Fig. 9 also demonstrate that tones interfere with one another and do so extensively. This kind of interference, however, does not seem to neutralize the tonal identities. For each tone, the greatest influence comes from the tone immediately preceding it, which greatly alters the onset and much of the early portion of its f_0 contour. The f_0 curves in the HLHLH sequence in Fig. 9, and many more in Figs. 4–6, if examined alone, could easily have been taken as instances of changed tonal identities. It was only when they were overlaid with f_0 curves that differed from them in just a single tone that it became apparent that their identities were not really lost (as shown in Figs. 4–6).

4.2.1. *The transition between two tones: How fast can it be?*

But why then should there be so much variation in the f_0 contour of a tone if its identity needs to be maintained? Possible answers may be found in the physiological constraints on pitch variation. When two linguistically specified pitch targets follow one another in an utterance, the vocal folds have to change their rate of vibration to make the pitch transition. A question then arises as to how fast this change can be. As discovered by Ohala and Ewan (1973), Ohala (1978), and Sundberg (1973, 1979), when asked to change pitch by 6 semitones in the shortest amount of time possible, a speaker needs at least 80–90 ms (female speakers being a bit faster) to complete 75% (i.e., the fastest central portion) of the change when raising the pitch, and 70–75 ms when lowering the pitch. (The exact mechanism that determines the speed of pitch change is still not very clear. See Fujisaki, 1988, and Titze, Jiang and Lin, 1997, for some discussion and speculation.) The duration for the complete change, though not reported in those studies, should be even longer. The average f_0 of the male voice in the present study is 117 Hz (obtained by taking a grand average over all the f_0 points in all 1920 f_0 curves produced by the four male speakers). At a center frequency of 117 Hz, six semitones corresponds to about 40 Hz, ranging from 97 to 137 Hz. This about covers the range of f_0 movement in a non-focused dynamic tone such as F or R. The average duration of non-focused syllables in the present study was found to be about 180 ms. This is barely enough time to complete two f_0 movements, each shifting 6 semitones. So, to produce a dynamic tone such as a F tone after a L tone, about half of the syllable duration would have to be used for the transition from the low f_0 offset of the preceding L tone to the high onset pitch of the falling F tone, while the other half has to be used for the falling contour itself, as indeed seems to be the case in panels C3R1-R3 of Fig. 4. It is thus not surprising to find a seemingly long f_0 transition between two tones whenever the pitch values differ substantially at the boundary, as can be seen clearly in Fig. 9.

4.2.2. *Where does the transition occur?*

If there has to be a transition between two tones, a further question is exactly where this transition may occur. As discussed earlier and as can be clearly seen in Fig. 9, the f_0 contour in the early portion of each syllable seems to be the most transitional: its starting

value varies extensively and almost exclusively with the ending f_0 of the preceding syllable. The f_0 trajectory then moves away gradually from the ending value of the previous tone and continues toward the typical contour shape of the current tone. In fact, the entire f_0 contour in any syllable seems to be a continuous transition away from the preceding tone as it moves ever closer to the designated tone contour of the current syllable, high for the H tone, rising for the R tone, low for the L tone, and falling for the F tone. This approaching movement is not completed until the end of the syllable. As soon as the syllable boundary is reached, however, another transition cycle begins, this time toward the target contour of the next tone. This transitional pattern has been reported in detail before (Xu, 1993, 1997). Similar patterns have also been found in Thai (Gandour *et al.*, 1994). It seems, therefore, that the transition toward a tone occurs *throughout* the duration of the syllable that carries the tone, although its rate of approximation slows down over time.

4.3. *How tones align with syllables*

The above discussion suggests that the syllable is the domain of tone implementation in Mandarin. The analysis of f_0 contour alignment discussed in 3.4.2 further examined how various critical points of a tonal contour are aligned to the syllable in Mandarin, and the results were summarized in 3.4.3. To interpret these alignment patterns, it is necessary to first understand the nature of the critical f_0 points (f_0 peaks and valleys) that were examined in the alignment analysis. First of all, a peak or a valley in an f_0 curve is also a turning point, i.e., a point at which the f_0 movement changes directions. Secondly, in a sentence without a narrow focus, a later f_0 peak can be observed in an H or R tone *only* when the following tone has a low onset pitch, such as in R or L. Likewise, a later f_0 valley can be observed in a F or L tone *only* when the following tone has a high onset pitch, such as in H or F. Assuming that both H and R tones have a late high pitch target, then a change of f_0 direction from going upward to downward may indicate that, at that moment, the implementation of the H or R tone is over and that of the next tone begins. The same interpretation is also applicable to the f_0 valleys in the L or F tone.

As shown in Table XIV and as can be seen in the HLHLH sequences in Fig. 9, in an H tone, if a later f_0 peak is discernable, it usually occurs right before the offset of the H-tone-carrying syllable. In addition, as described in 3.4.2.3, the distribution of the peak location as a function of syllable duration in the H tone is similar to cloud 3 in Fig. 8, indicating that the f_0 peak moves more in synchrony with the syllable offset than with the onset. Also as shown in Fig. 9 and discussed earlier, it is near the end of the syllable that a tonal target is most closely approximated. This means that it is around the moment when the target contour of the H tone is best approximated that the moment toward the following tone begins. Since that moment remains close to the offset of the H-tone-carrying syllable as revealed by the alignment analysis, it seems that the boundary between two syllables is probably used as a reference point for the timing of tone implementation. At the same time, the fact that the f_0 peak in the H tone occurs before syllable offset indicates that there is probably enough time for the high pitch target of the H tone to be reached before the end of the syllable. Also as discussed in 3.4.2.3, the alignment of the later f_0 valley in the L tone has a similar pattern, occurring before and remaining close to the syllable offset.

In contrast to the H tone, the f_0 peak in the R tone, when observable (i.e., before an R or L tone), usually occurs *after* the offset of the R-tone-carrying syllable (while

remaining close to it). The difference between the H and R tones in terms of peak alignment is an interesting one. Two possible accounts may be considered. First, it may be the case that the R tone has two successive pitch targets—low + high. Due to the limit on the rate of pitch change, there may not be enough time to realize the final high pitch target by the end of the syllable, hence a "spill over" (Ohala, 1973:31) results. In contrast, the limit on the rate of pitch change should not be a problem for the H tone because for a duration of around 199 ms (cf. Table XIV), there should be plenty of time to reach a single high pitch target. Alternatively, it may be the case that the R tone in Mandarin is intrinsically dynamic, while the H tone is intrinsically static. As discussed earlier, a tonal contour is maximally approximated by the end of the syllable. Hence, the dynamic movement of the R tone should be most fully realized by the end of the syllable. Just as a pitch change takes time to implement, a change of f_0 movement direction cannot happen instantaneously either. If the R tone is indeed so implemented that the f_0 curve keeps going up until the syllable offset, the directional change has to occur somewhere after the syllable offset. In physiological terms, if the articulatory movement producing the f_0 rise continues until the end of the syllable, the movement cannot be terminated immediately, but only a short while after the syllable offset. Hence, if the R tone is intrinsically dynamic, it would also be natural that f_0 did not start descending right at syllable offset when the R tone is followed by a tone with a low beginning target.

In addition to the later extreme points, sometimes an earlier extreme f_0 point can be also observed. When a F tone is preceded by a tone with a low final pitch, there is usually an f_0 peak in the F-tone-carrying syllable, as can be seen in panels C3R1-4 in Fig. 4. This peak, however, should not be viewed as the onset of the F tone. This is because to reach the initial high f_0 of the F tone from the low f_0 of the preceding tone, there has to be an initial transition, and this transition has been found to start near the syllable onset, as discussed above. So, the location of an earlier f_0 peak in a F tone may be interpreted more appropriately as the moment at which the falling contour of the F tone starts to be *effectively* realized. This moment apparently differs depending on how much distance in pitch the initial transition has to cover, as can be seen in panels C3R1-4 of Fig. 4 — later in the LF sequence because of the large pitch difference, but earlier in the FF sequence because of the smaller pitch difference. This agrees with what can be predicted by the limit on the rate of pitch change. However, if the magnitude of the initial transition and rate of pitch change are the only two factors determining the location of the early f_0 peak in the F tone, when the syllable duration increases, the f_0 peak should be reached relatively earlier in a syllable than when duration is shorter. As discussed in 3.4.2.2, however, just the opposite is the case: the longer the syllable duration, the later the relative location of the f_0 peak in the F-tone-carrying syllable. It seems that when syllable duration is increased, there is an effort not to reach the f_0 peak too early. Since an earlier peak would have reduced the slope of the falling contour, the effort is probably to maintain the falling rate in the F tone. A similar tendency was also found with the R tone, as discussed in 3.4.2.2. Relating this tendency to the unresolved issue regarding the nature of peak delay in the R tone, it seems that it is probably due to the maintenance of the rising contour as a dynamic movement that the f_0 peak is delayed beyond the offset of the R-tone-carrying syllable.

To sum up, the results of f_0 contour alignment analysis indicate that a tone in Mandarin is probably implemented synchronously with the entire syllable that carries it. However, there is also a tendency to postpone the maximum approximation of a dynamic

tone (R or F) to the later portion of a syllable when the syllable duration is long in order to maintain the integrity of the tonal contour.

These alignment patterns have recently been further confirmed in a study (Xu, 1998) that examined tonal alignment in syllables with a final nasal and in syllables with an initial voiceless consonant. It thus seems that the alignment patterns found in the present study remain consistent across different syllable structures in Mandarin.

4.4. *How tone and focus affect more global f_0 contours*

Having discussed how tone and focus are implemented in Mandarin, it is now possible to address the last major question of the present study, namely, how tone and focus jointly determine the formation of surface f_0 contours of an utterance. Again, much of the following discussion will refer to the mean f_0 tracings shown in Fig. 9.

4.4.1. *Downtrends due to tonal interaction*

As demonstrated by Table VIII and illustrated by the upper left panel of Fig. 9, in the short sentences examined in the present study, when all the syllables carry the H tone, little overall downtrend is evident (about 2.5 Hz between the first and last H tones in Fig. 9). Such weak declination is in sharp contrast with the declination rates of -20–30 Hz/s as suggested by Maeda (1976), -11 semitones/$(t + 1.5)$ for $t \leq 5$ s as suggested by 't Hart (1979), and -10 Hz/s as suggested by Pierrehumbert and Beckman (1988). When the H tones are interrupted by the L tone, however, an overall downward tilt becomes clearly visible. As can be seen in Fig. 9, this downward tilt is likely due to two relatively local effects: carryover lowering and anticipatory raising. Carryover lowering is mostly due to the f_0 transition between two adjacent tones. Anticipatory raising, whose underlying mechanism is still unclear, has a dissimilatory effects on the f_0 of the preceding tone: the lower the minimum f_0 of the following tone, the higher the maximum f_0 of the preceding tone. When the effects of carryover lowering and anticipatory raising are combined, the f_0 of the first H in a HLHLH sequence is much higher than that of the last H. In the upper left panel of Fig. 9, the first H is 10 Hz higher than the second H; and the second H is 17 Hz higher than the last H (disregarding the initial sharp f_0 drop due to the initial voiceless stop in syllable 5). The two differences combined then gives a downward tilt of greater than 27 Hz/s (with the mean duration of these utterances less than a second), which is more like the declination rates suggested by Maeda (1976).

4.4.2. *Downtrends related to focus and other factors*

Also shown in Table VIII and illustrated by Fig. 9 is that focus can also bring about substantial downtrends. In fact, a non-final focus seems to have greater lowering power than any of the lexical tones (rows 2 and 3 in Table VIII), and it can also boost the lowering power of the tones under focus, although not always consistently (e.g., for the L tone, as discussed earlier). In the lower left panel of Fig. 9, a narrow focus on word 1 introduces a difference of 59 Hz (9 semitones) between the second and last H tones in the HHHHH tone sequence, which is almost as large as the declination rate suggested by 't Hart (1979).

In addition to tone and focus, there may be other factors that can also bring about additional downtrend in an utterance. A recent study of Mandarin by Shih (1997) reported a major downtrend in sentences consisting of all H tones except for the first two. A comparison of Shih (1997) and the present study found several differences that may

account for the discrepancy between the findings. First, sentences examined in Shih (1997) varied in length (4–13 syllables), and as she reported, shorter sentences had lower initial H tone and higher final H tone. Second, all sentences examined by Shih had a LR tone sequence in the first word. Finally, Shih described the declination pattern over the course of the sentence as non-linear and exponential, comparable to the exponential downstep model proposed by Pierrehumbert (1980). A similar nonlinear decline in f_0 was also reported by Gelfer, Harris, Collier and Baer (1985) for Dutch and by Prieto *et al.* (1996) for Mexican Spanish. It has been proposed by Umeda (1982) that an exceedingly high f_0 peak at the onset of the first sentence of a paragraph is probably used as a beginning signal for new topics, or is so produced to draw the listeners' attention. Based on this proposal, the exponential decay of peak f_0 values over the course of a sentence can be interpreted as partly due to a very high initial f_0 peak followed by a drop to a relatively neutral pitch level. (See also Nakajima & Allen, 1993, for convincing data on topic-initiation related initial high f_0 values.) In the present study, all sentences, including those with neutral focus, were elicited as answers to different questions. In Shih (1997), in contrast, only sentences with focus on the first word were elicited as answers to specific questions, whereas those without focus were produced as read statements. It could be the case that when read as isolated statements, the sentences may be implemented as each introducing a new topic. If so, topic initiation may introduce more f_0 decline in addition to that due to focus and tonal interactions, thus bringing the rate of overall f_0 decline closer to or even greater than the declination rate suggested by 't Hart (1979). This is certainly an area where further investigation is needed.

4.4.3. *Transitions due to continuity of f_0 contours*

The mean f_0 tracings in Fig. 9 also illustrate the effect of a factor that plays an important role in the formation of surface f_0 contours. That is, f_0 production seems to be *continuous* as long as voicing is not interrupted. Because of such continuity, when two tones are produced next to each other, there has to be a transition between them, and because of the limit on the rate of pitch change, such a transition may appear as an apparent rise or fall. Thus, all the L tones in Fig. 9 appear to have a falling contour, because to reach their low target they have to travel from the previous H tone. Likewise, the second H tones in the HLHLH sequences in Fig. 9 all have an apparent rising contour occupying much of their duration, and this seems to be due to the fact that they all have to travel from the preceding L tone to reach the high target. Also as demonstrated by Figs. 4–6, when a dynamic tone such as R or F is produced next to another tone, there again has to be a transition between the two tones, which may appear as either rising, falling, or flat, depending on the nature of the f_0 difference at the junction. For example, a tone sequence of RR entails an intervening falling transition if voicing continues through the syllable boundary. That falling transition should not be considered as an inserted F tone unless there is independent evidence for the occurrence of such epenthesis. In all these cases, therefore, it is important to recognize from the surface f_0 curve which regions correspond directly to the real tonal targets and which are merely transitions between them.

The transition between two tones (or at least the most conspicuous portion of it) may be obliterated by an intervening obstruent consonant, however, because an obstruent may not only interrupt voicing, but also change the f_0 height of the following vowel, as has been well established (Lehiste & Peterson, 1961; Hombert, 1978; Santen &

Hirschberg, 1994). A voiceless obstruent may significantly raise the portion of the f_0 contour immediately following it (Hombert, 1978; Santen & Hirschberg, 1994), thus significantly reducing a rising f_0 transition, as can be seen in the last L tone in the HLHLH sequences in Fig. 9. Previous studies have also reported that in certain tone languages, an obstruent may block a phonological process known as tone spreading—a left-to-right tonal assimilation (Hyman & Schuh, 1974; Schuh, 1978). As described by Schuh (1978), a left-to-right H-tone spreading may be blocked by a voiced obstruent, and left-to-right L-tone spreading may be blocked by a voiceless obstruent. There is, therefore, an interesting parallel between the characteristics of these blocking rules and the known effects of obstruents on f_0. That is, voiceless obstruents that may raise f_0 might also block the spreading of a L into the following syllable; and voiced obstruents that may lower f_0 might also block the spreading of a H tone. It could be the case, therefore, that what is actually blocked is the f_0 transition between two adjacent tones rather than the spreading itself.

An apparent f_0 transition may also occur whenever there is a difference in f_0 between two adjacent syllables, regardless of whether or not the two syllables carry the same tone. In the lower left panel of Fig. 9, for example, there is an apparent falling transition between the second and third H tones in the HHHHH sequence. Similarly, there is an apparent rising transition between the second and third H tones in the HHHHH sequence in the lower right panel of Fig. 9. These transitions are between syllables with the same tone that is differently affected by the focus. They do not constitute a change of the underlying tones, because they occur in the earlier rather than the later portion of the syllable.

Finally, as can be seen in all the HLHLH sequences in Fig. 9, the maximum f_0 in the first syllable is increased due to anticipatory raising. However, the maximum f_0 is not reached until near the end of the syllable. In addition, when the narrow focus is on the first word (which is disyllabic) in the HHHHH sequence (lower left panel), the f_0 height, though already greater than other focus conditions from the very beginning, continues to rise until near the end of the focused word. The rising contours in both cases resemble the onset ramp reported for many other languages. Vaissière (1995) even considers this onset ramp as an intonation universal. As is apparent in Fig. 9, for these Mandarin utterances, at least, this kind of onset ramp is merely a transition between the f_0 onset of the utterance and the first f_0 maximum, which does not seem to constitute a functional rise.

5. Conclusions

The results of the present study reveal that the formation of f_0-contours in short declarative sentences in Mandarin can be mostly explained in terms of the contribution of lexical tone and focus. These contributions were found to determine much of the global shape, the local contours, the slope of rising and falling curves, and the alignment of the contours with the syllables. Focus was found to significantly influence the global f_0 shape of the entire sentence. A non-final focus substantially expands the pitch range, particularly the upper end, of the words directly under focus, and it suppresses the pitch range of post-focus words, while leaving that of pre-focus words largely intact. This robust asymmetry about the focus was found to generate a substantial global decline in f_0 over the course of an utterance whenever a non-final focus occurred.

More locally, a lexical tone was found not only to determine the shape of the f_0 contour of the syllable that carries it, but also to influence the shape and height of the

f_0 contours in surrounding syllables. Its influence is assimilatory and substantial on the following tones, but dissimilatory and relatively subtle on the preceding tone. Under the heavy carryover influence, a tone is realized by making a continuous transition from the ending f_0 of the preceding tone to the current target contour. This transition seems to start from the onset of the syllable that carries the tone and continue throughout the syllable. As a result, the proper contour of a tone is most closely approximated in the later portion of a syllable, while the influence of the preceding tone appears most salient in the earlier portion of the syllable.

The f_0 peak associated with the R tone and the f_0 valley associated with the F tone, though generally occurring *after* syllable offset, were found to remain close to and move highly in synchrony with syllable offset. The f_0 peak associated with the H tone and f_0 valley associated with the L tone were found to generally occur *before* syllable offset but also remain close to and move mostly in synchrony with syllable offset. The earlier extreme points in the R and F tones — f_0 valley in the R tone and f_0 peak in the F tone — were found to occur near the center of the syllable but closer to syllable offset than the onset. They were also found to move somewhat more in synchrony with syllable offset than the onset. These alignment patterns were interpreted as further evidence that tones in Mandarin are implemented synchronously with the associated syllables and realized in such a way that their target contours are best approximated by the end of the syllable.

It was also found that the interaction between adjacent tones in the form of carryover lowering and anticipatory raising often generated an f_0 decline over time when there was a non-H tone in the utterance. The L tone was found to generate a greater decline than the R and F tones, and the magnitude of such declines was found to increase with the number of non-H tones in the utterance. An early focus and a non-H tone, therefore, both generated f_0 downtrends. When the two effects were combined, substantial f_0 decline over the course of an utterance usually resulted. It was suggested, therefore, that these may be two of the important mechanisms (in addition to topic initiation, which was apparently absent in the utterances examined in the present study) underlying the f_0 phenomena known as downstep and declination.

Finally, it is particularly worth mentioning that Mandarin is a language whose tonal space is relatively crowded—with four lexical tones, two of which have dynamic contours. It therefore may not be a language that would allow for the maximum amount of carryover lowering, anticipatory raising, or focus-related f_0 variations. Nonetheless, substantial variations in the shape and height of f_0 contours due to these effects were found in Mandarin. It is therefore possible that similar effects with greater magnitude than have been found for Mandarin may occur in languages whose tonal space is less crowded, i.e., with fewer or no lexical tones. Further studies are needed to verify this possibility.

I would like to thank Ignatius G. Mattingly, D. H. Whalen, and Bruce Smith for valuable comments on an earlier version of this paper. I am also grateful to Mary Beckman, Rebecca Herman, and an anonymous reviewer for their helpful comments and suggestions.

References

Abramson, A. S. (1962) *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments.* Bloomington: Indiana University Research Center in Anthropology, Folklore, and Lingusitics, Pub. 20.

Abramson, A. S. (1976) Thai tones as a reference system. In *Thai linguistics in honor of Fang-Kuei Li* (T. W. Gething, J. G. Harris & P. Kullavanijaya, editors), pp. 1–12. Bangkok: Chulalongkorn University Press.
Abramson, A. S. (1978) Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, **21**, 319–325.
Arvaniti, A. & Ladd, D. R. (1995) Tonal alignment and the representation of accentual targets. In *Proceedings of The 13th International Congress of Phonetic Sciences*, Stockholm, **4**, pp. 220–223.
Arvaniti, A., Ladd, D. R. & Mennen, I. (1998) Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, **36**, 3–25.
Bai, D. (1934) Guanzhong shengdiao shiyan lu [Experiments with tones of Guanzhong dialects]. In *Shiyusuo Jikan [A Collection by Shiyusuo]* pp. 355–361.
Beckman, M. E. (1995) Local shapes and global trends. In *Proceedings of The 13th International Congress of Phonetic Sciences*, Stockholm, **2**, pp. 100–107.
Bunnell, H. T., Hoskin, S. R. & Yarrington, D. (1997) Interactions among f_0, duration, and amplitude in the perception of focus, *Journal of the Acoustical Society of America*. **102, Pt 2**, pp. 3203–3204.
Chao, Y. R. (1948) *Mandarin Primer*. Cambridge: Harvard University Press.
Chao, Y. R. (1956) Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In *For Roman Jakobason* (M. Halle, editor), pp. 52–59 Mouton: The Hague.
Chao, Y. R. (1968) *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.
Chuang, C. K., Hiki, S., Sone, T. & Nimura, T. (1971) The acoustical features and perceptual clues of the four tones of standard colloquial Chinese. In *Proceedings of The Seventh International Congress on Acoustics*, Budapest, **25 C 13**, pp. 297–300.
Cohen, A., Collier, R. & 't Hart, J. (1982) Declination: Construct or intrinsic feature of spech pitch, *Phonetica*, **39**, 254–273.
Chen, A. & 't Hart, J. (1965) Perceptual analysis of information patterns. In *Proceedings of the Fifth International Congress on Acoustics* (D.E. Commins, editor), pp. A. 16. Liège.
Collier, R. (1975) Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America*, **58**, 249–255.
Collier, R. (1984) Some physiological and perceptual constraints on tonal systems. In *Explanations for language universals* (B. Butterworth, B. Comrie & O. Dahl, editors), p. 237–248. Amsterdam: Mouton.
Collier, R. (1987) F0 declination: The control of its setting, resetting, and slope. In *Laryngeal Function in Phonation and Respiration* (T. Baer, C. T. Sasaki & K. S. Harris, editors). pp. 403–421. Boston: College-Hill Press.
Cooper, W. E., Eady, S. J. & Mueller, P. R. (1985) Acoustical aspects of contrastive stress in question-answer contexts, *Journal of the Acoustical Society of America*, **77**, 2142–2156.
Cooper, W. E. & Sorensen, J. M. (1977) Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America*, **62**, 683–692.
Cooper, W. E. & Sorensen, J. M. (1981) *Fundamental frequency in sentence production*. Springer-Verlag: New York.
Eady, S. J. & Cooper, W. E. (1986) Speech intonation and focus location in matched statements and questions, *Journal of the Acoustical Society of America*, **80**, 402–416.
Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R. & Lotts, D. W. (1986) Acoustic characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments, *Language and Speech*, **29**, 233–251.
Fujisaki, H. (1988) A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour, In *Vocal Physiology: Voice Production*, (O. Fujimura, editor), pp. 347–355. New York: Raven Press, Ltd.
Gandour, J., Potisuk, S. & Dechongkit, S. (1994) Tonal coarticulation in Thai, *Journal of Phonetics*, **22**, 477–492.
Gandour, J., Potisuk, S., Dechongkit, S. & Ponglorpisit, S. (1992) Anticipatory tonal coarticulation in Thai noun compounds, *Linguistics of the Tibeto-Burman Area*, **15**, 111–124.
Gårding, E. (1987) Speech act and tonal pattern in Standard Chinese, *Phonetica*. **44**, 13–29.
Gelfer, C. E., Harris, K. S., Collier, R. & Baer, T. (1985) Is declination actively controlled? In *Vocal Fold Physiology: Biomechanics and Phonatory Control* (I. R. Titze & R. C. Scherer, editors), pp. 113–126. Denver, CO: Denver Center for the Performing Arts.
Han, M. S. & K.-O. Kim (1974) Phonetic variation of Vietnamese tones in disyllabic utterances, *Journal of Phonetics*, **2**, 223–232.
Hermes, D. J. (1997) Timing of pitch movements and accentuation of syllables in Dutch. *Journal of the Acoustical Society of America*, **102**, 2390–2402.
Ho, A. T. (1976) Mandarin tones in relation to sentence intonation and grammatical structure. *Journal of Chinese Linguistics*, **4**, 1–13.
Hombert, J.-M. (1978) Consonant types, vowel quality, and tone. In *Tone: A linguistic survey* (V. A. Fromkin, editor), pp. 77–111. New York: Academic Press.
Howie, J. M. (1974) On the domain of tone in Mandarin, *Phonetica*, **30**, 129–148.
Howie, J. M. (1976) *Acoustical Studies of Mandarin Vowels and Tones*. London: Cambridge University Press.

Hyman, L. M. (1973) The role of consonant types in natural tonal assimilations. In *Consonant Types and Tone* (L. M. Hyman, editor), pp. 151–179. Los Angeles, CA: Department of Linguistics, University of Southern California.
Hyman, L. M. (1993) Register tones and tonal geometry. In *The Phonology of Tone* (H.v.d. Hulst & K. Snider, editors), pp. 75–108. New York: Mouton de Gruyter.
Hyman, L. & R. Schuh (1974) Universals of tone rules. *Linguistic Inquiry*, **5**, 81–115.
Jin, S. (1996) *An Acoustic Study of Sentence Stress in Mandarin Chinese*. Ph.D. dissertation, The Ohio State University.
Kim, S.-A. (in press) Positional effect on tonal alternation in Chichewa: Phonological rule vs. phonetic timing. In *Proceedings of Chicago Linguistic Society*, 34.
Laniran, Y. (1992) *Intonation in Tone Languages: the phonetic implementation of tones in Yorùbá*. Unpublished Ph.D. dissertation, Cornell University.
Laniran, Y. O. & Clements, G. N. (1995) A long-distance dependency in Yorùbá tone realization. In *Proceedings of The 13th International Congress of Phonetic Sciences*. Stockholm, **2**, pp. 734–737.
Laniran, Y. & Gerfen, C. (1997) High raising, downstep and downdrift in Igbo. In *Proceedings of The 71st Annual Meeting of the Linguistic Society of America*, Chicago, p. 59.
Lehiste, I. (1970) *Suprasegmentals*. Cambridge. MA: MIT Press.
Lehiste, I. & Peterson, G. E. (1961) Some basic considerations in the analysis of intonation, *Journal of the Acoustical Society of America*, **33**, 419–425.
Liberman, M. & Pierrehumbert, J. (1984) Intonational invariance under changes in pitch range and length. In *Language Sound Structure* (M. Aronoff & R. Oehrle, editors), pp. 157–233. Cambridge, Massachusetts: M.I.T. Press.
Lieberman, P. (1967) *Intonation, perception and language*. Cambridge, MA: MIT Press.
Lin, M.-C. (1965) Yingao xianshiqi yu Putonghu shengdiao yingao texing [The pitch indicator and the pitch characteristics of tones in Standard Chinese]. *Zhongguo Yuwen [Chinese Lingusitics]*, **204**, 182–93.
Lin, M.-C. (1988) Putonghua shengdiao de shengxue texing he zhijue zhengzhao [The acoustic characteristics and perceptual cues of tones in Standard Chinese]. *Zhongguo Yuwen [Chinese Linguistics]*, **204**, 182–193.
Lin, M. & Yan, J. (1991) Tonal coarticulation patterns in quadrisyllabic words and phrases of Mandarin. In *Proceedings of The 12th International Congress of Phonetic Sciences*. **3**, pp. 242–245.
Lorch, R. F. & Myers, J. L. (1990) Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16**, 149–157.
Maeda, S. (1976) *A Characterization of American English Intonation*. Cambridge, MA: MIT Press.
Meeussen, A. E. (1970) Tone typologies for West African Languages. *African Languages Studies*, **11**, 266–71.
Morton, J., Marcus, S. & Frankish, C. (1976) Perceptual centers (P-centers). *Psychological Review*, **83**, 405–408.
Nakajima, S. & J. F. Allen (1993) A study on prosody and discourse structure in cooperative dialogue, *Phonetica*, **50**, 197–210.
Ohala, J. J. (1973) The physiology of tone. In *Consonant Types and Tone* (L. M. Hyman editor), pp. 1–14. Los Angeles, CA: Department of Linguistics, University of Southern California.
Ohala, J. J. (1978) The production of tone. In *Tone: A linguistic survey* (V. A. Fromkin, editor), pp. 5–39, New York: Academic Press.
Ohala, J. J. (1990) Respiratory activity in speech. In *Speech production and speech modelling* (Hardcastle & Marchal, editors), pp. 23–53. Dordrecht: Kluwer.
Ohala, J. J. & Ewan, W. G. (1973) Speed of pitch change. *Journal of the Acoustical Soiety of America*, **53**, 345(A).
Pierrehumbert, J. (1979) The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, **66**, 363–369.
Pierrehumbert, J. (1980) *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, Massachusetts Institute of Technology.
Pierrehumbert, J. & Beckman, M. (1988) *Japanese Tone Structure*. Cambridge, MA: MIT Press.
Pike, K. L. (1945) *The Intonation of American English*. Ann Arbor: University of Michigan Press.
Pike, K. L. (1948) *Tone Languages*. Ann Arbor: University of Michigan Press.
Poser, W. (1984) *The phonetics and phonology of tone and intonation in Japanese*. Ph.D. dissertation, MIT, Cambridge. MA.
Prieto, P., Santen, J. v. & Hirschberg, J. (1995) Tonal alignment patterns in Spanish, *Journal of Phonetics*, **23**, 429–451.
Prieto, P., Shih, C. & Nibert, H. (1996) Pitch downtrend in Spanish. *Journal of Phonetics*, **24**, 445–473.
Rose, P. J. (1998) On the non-equivalence of fundamental frequency and pitch in tonal description. In *Prosodic Analysis and Asian Linguistics: To Honour R. K. Sprigg* (D. Bradley, E. J. A. Henderson & M. Mazaudon, editors), pp. 55–82. Canberra: Pacific Linguistics.
Santen, J. P. H. v. & Hirschberg, J. (1994) Segmental effects on timing and height of pitch contours. In *Proceedings of The International Conference on Spoken Language Processing*, **94**, pp. 719–722.
Schuh, R. G. (1978) Tone Rules. In *Tone: A linguistic survey* (V. A. Fromkin, editor), pp. 221–256. New York: Academic Press.

Shen, J. (1994) Hanyu yudiao gouzao he yudiao leixing [Intonation structures and patterns in Mandarin], *Zhongguo Yuwen* [*Journal of Chinese Linguistics*], **1994-3,** 221–228.
Shih, C.-L. (1988) Tone and intonation in Mandarin, *Working Papers, Cornell Phonetics Laboratory*, **No. 3,** 83–109.
Shih, C.-L. (1997) Declination in Mandarin. In *Intonation: Theory, Models and Applications, Proceedings of an ESCA Workshop. European Speech Communication Association* (A. Botinis, G. Kouroupetroglou & G. Carayannis, editors), pp. 293–296. Athens, Greece: European Speech Communication Association.
Silverman, K. E. A. & Pierrehumbert, J. B. (1990) The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology 1 – Between the Grammar and Physics of Speech* (J. Kington & M. E. Beckman, editors), pp. 72–106. Cambridge: Cambridge University Press.
Steele, S. A. (1986) Nuclear accent f_0 peak location: Effects of rate, vowel, and number of following syllables, *Journal of the Acoustical Society of America*, **80,** S51.
Stewart, J. M. (1965) *The typology of the Twi tone system.* Legon, Ghana: Institute of African Studies, University of Ghana.
Stewart, J. M. (1983) Key lowering (downstep/downglide) in Dschang, *Journal of African Languages and Linguistics*, **3,** 113–138.
Sundberg, J. (1973) Data on maximum speed of pitch changes, *STL-QPSR.* **4,** 39–47.
Sundberg, J. (1979) Maximum speed of pitch changes in singers and untrained subjects, *Journal of Phonetics*, **7,** 71–79.
Titze, I. R. & Durham, P. L. (1987) Passive mechanisms influencing fundamental frequency control. In *Laryngeal Function in Phonation and Respiration* (T. Baer, C. Sasaki & K. S. Harris, editors), pp. 304–319. Boston: College-Hill Press.
Titze, I. R., Jiang, J. J. & Lin, E. (1997) The dynamics of length change in canine vocal folds, *Journal of Voice*, **11,** p. 267.
't Hart, J. (1979) Explorations in automatic stylization of f_0 curves, *IPO Annual Progress Report*, **14,** 61–65.
't Hart, J. & Collier, R. (1975) Integrating different levels of intonation analysis. *Journal of Phonetics*, **3,** 235–256.
Umeda, N. (1982) "f_0 declination" is situation dependent, *Journal of Phonetics*, **10,** 279–290.
Vaissière, J. (1995) Phonetic explanations for cross-linguistic prosodic similarities, *Phonetica*, **52,** 123–130.
Wang, W. S.-Y. & Li, K.-P. (1967) Tone 3 in Pekinese, *Journal of Speech and Hearing Research*, **10,** 629–636.
Whalen, D. H. & Xu, Y. (1992) Information for Mandarin tones in the amplitude contour and in brief segments, *Phonetica*, **49,** 25–47.
Xu, Y. (1993) *Contextural Tonal Variation in Mandarin Chinese*, Ph.D. dissertation. The University of Connecticut.
Xu, Y. (1994) Production and perception of coarticulated tones, *Journal of the Acoustical Society of America*, **95,** 2240–2253.
Xu, Y. (1995) The effect of emphatic accent on contextual tonal variation. In *Proceedings of The 13th International Congress of Phonetic Sciences*, Stockholm, **3,** pp. 668–671.
Xu, Y (1997) Contextual tonal variations in Mandarin, *Journal of Phonetics*, **25,** 61–83.
Xu, Y. (1998) Consistency of tone-syllable alignment across different syllable structures and speaking rates, *Phonetica*, **55,** 179–203.
Xu, Y. & Kim, J. (1996) Downstep, regressive upstep, H-raising, or what? — Sorting out the phonetic mechanisms of a set of "phonological" phenomena, *Journal of the Acoustical Society of America*, **100, Pt 2,** p. 2824.

Appendix 1

Description of the C code for the trimming algorithm used to smooth raw f_0 curves: The trimming algorithm compared three f_0 points at a time. If the middle point is greater than (or smaller than) both flanking points by the amount specified by MAXBUMP and MAXEDGE, it is replaced by a point that makes the line between the flanking points a straight one. This trimming algorithm effectively eliminates sharp spikes in the raw f_0 tracing often seen around nasal-vowel junctions. In contrast, the triangular smoothing algorithm commonly used would always retain some effects of the spike, since its value is included in the running means. This is particularly critical for the f_0 peak measurements taken in the present study. Even when a small effect of the spike is left in the curve, the smoothed bump at that location could still be taken as an f_0 peak by an automatic peak searching algorithm. The C code for the algorithm can be obtained by contacting the author.