Original Paper

Phonetica 1998;55:179–203

Received: April 18, 1998
Accepted: August 21, 1998

Consistency of Tone-Syllable Alignment across Different Syllable Structures and Speaking Rates

Yi Xu

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

Abstract

Previous studies have reported that F_0 contours of lexical tones in Mandarin are closely aligned with the syllables that carry them and that this alignment is late-adjusted so that the most appropriate F_0 contour of a tone occurs mainly in the later portion of the host syllable. These studies, however, examined tone alignment only in syllables with a simple CV structure. The present study compared tone-syllable alignment patterns in syllables with a final nasal (the only final consonant allowed in Mandarin) to those without a final nasal at three speaking rates – normal, fast and slow. Experiment 1 found that regardless of internal syllable structure, the F_0 contours for all of the four Mandarin tones maintain a consistent alignment to the syllables that carry them. Experiment 2 found that whether or not the syllable has a final nasal and regardless of speaking rate, the F_0 peak associated with the R (rising) tone always occurs near the offset of the R-tone-carrying syllable, the onset of the F_0 rise always occurs near the center of the syllable and the slope of the rise does not vary systematically with either syllable structure or speaking rate. These findings are interpreted as indicating that the syllable is the proper domain for tone implementation and that tone contours such as rising and falling are probably implemented as integral dynamic targets rather than as sequences of static targets.

Introduction

How F_0 contours align with other units in speech is an issue that is important not only for modeling F_0 contours in speech synthesis and in processing F_0 contours for speech recognition but also for theoretical understanding of tone and intonation in general. The alignment issue has received increasing attention from researchers of tone and intonation in recent years. There have been several lines of research investigating F_0 contours alignment from different perspectives. The first line of research concerns how lexical tones are aligned to other units in speech. In languages such as Mandarin and Thai, every syllable is assigned a tone in the lexicon. Because of that, it

KARGER

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

0031–8388/98/0554–0179
$17.50/0
Article accessible online at:
http://BioMedNet.com/karger

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

is generally assumed that tones are associated and hence aligned with the syllable in those languages. At a more phonetic level, however, there have been questions as to whether a tone is carried by the entire syllable or by only a portion of the syllable. Howie [1974] reports that tones in Mandarin are carried by only the syllable rhyme, while the portion of the F_0 contour corresponding to an initial voiced consonant or glide is merely an adjustment for the voicing of initial consonants. More recently, however, Lin [1995] argued that neither initial consonants and glides nor final nasals play any tone-carrying role in Mandarin. Apparently, more research is needed to examine the detailed tonal alignment patterns in various tone languages, including Mandarin.

The second line of research is concerned with how F_0 peaks align with the associated syllables in languages without lexical tones. There have been several studies that examine this issue, but the findings have been somewhat diverse. To investigate F_0 peak alignment, a number of studies have used a measurement called 'peak delay', which is the distance between the F_0 peak and the onset of either the syllable or the nuclear vowel of the syllable associated with the peak. This peak delay measurement is then further examined as a function of either vowel duration [Steele, 1986; Silverman and Pierrehumbert, 1990] or syllable duration [Prieto et al., 1995]. Xu [forthcoming] reevaluated how these peak delay data could be interpreted and observed that syllable duration was actually equivalent to the location of syllable offset relative to the syllable onset and vowel duration (in a syllable without final consonant) equivalent to the location of syllable offset relative to the vowel onset. It was thus argued that the regression of 'peak delay' to syllable or vowel duration in fact reveals the timing relationship between the relative locations of the F_0 peaks and the syllable offset: a regression coefficient of 1 (combined with an intercept of zero) indicates that the F_0 peak tends to occur near the syllable offset, whereas a regression value smaller than 1 indicates that the F_0 peak tends to lag behind the syllable offset as the syllable or vowel duration increases. Applying this interpretation to Mandarin, Xu [forthcoming] found that F_0 peaks associated with the rising (R) tone always occurred immediately after the syllable offset; F_0 peaks associated with the high (H) tone occurred somewhere before the syllable offset but moved rightward with the syllable offset as syllable duration increased; even F_0 peaks associated with the falling (F) tone occurred in the later half of the host syllables and moved more in synchrony with the syllable offset than with the syllable onset.

The third line of research related to F_0 contour alignment concerns whether all phonetically dynamic F_0 contours are composed of simple underlying pitch targets such as H and L or, alternatively, dynamic targets such as R (rise) and F (fall) also exist as primary pitch targets. The more traditional views of intonation describe intonation patterns as consisting of both pitch levels and pitch contours such as rise and fall [Bolinger, 1951; O'Connor and Arnold, 1961; 't Hart and Cohen, 1973; Clark, 1978; Ladd, 1978]. In contrast, many other intonation researchers argue that simple pitch targets such as H and L are the most basic components of intonation, and dynamic contours such as rise and fall are derived from concatenated static pitch targets, i.e. rise = LH and fall = HL. The most fully developed of such theories are presented by Pierrehumbert [1980], Liberman and Pierrehumbert [1984] and Pierrehumbert and Beckman [1988]. A similar debate has been going on concerning lexical tones. While some studies [Pike, 1948; Wang, 1967; Abramson, 1978] argue that contour tones found in languages such as Thai and Mandarin should be considered as single units, others [Woo, 1969; Leben, 1973; Gandour, 1974; Duanmu, 1994] treat contour tones as sequences of H and L targets. (A detailed account of the early debate can be found in Anderson [1978].) What might be helpful for settling this kind of debate is to examine how F_0 contour patterns change as a function of various factors, including tonal context, degree of stress, speaking rate, syllable structure, etc. The first two of these factors have been examined recently [Xu, forthcoming], and evidence has been found suggesting the possible existence of dynamic contours as primary pitch targets, but the issue is far from being settled.

The fourth line of research concerns how F_0 contours are perceptually aligned to segments, because perception reflects the listener's understanding of the speaker's production. House [1985, 1990] examined the perceptual alignment of F_0 contours and arrived at a tone perception model. He proposed that there are four basic patterns of perceived tones, namely H, L, R and F. He further proposed that one of the major determinants of tone perception is the alignment of the F_0 contour relative to the region of rapid spectral change. A dynamic (R or F) tone is perceived only if the dynamic portion of the F_0 contour occurs within a region with no rapid spectral change. If the dynamic portion of the F_0 contour occurs across a region of rapid spectral change, it will be heard as two static (H or L) tones. The regions of rapid spectral change that House examined, however, all correspond to initial consonants, such as /m/ in /ma/. This leaves open the question of how an F_0 contour corresponding to a final sonorant consonant, such /m/ in /am/ is perceived. His model would predict that a dynamic F_0 contour occurring across the vowel-sonorant boundary would not be heard as a dynamic tone. If this is the case, to produce an R tone in a syllable with a final sonorant, speakers would have to implement all (or at least the most) dynamic portion of the rising contour inside the vowel rather than over the entire rhyme. If, however, the syllable boundaries rather than regions of rapid spectral change are the reference points used by both listeners and speakers, the most critical portion of an F_0 contour would occur across the vowel-sonorant boundary within a syllable despite the rapid spectral change at that boundary.

Finally, our own research in the past few years has revealed interesting evidence as to why tone-associated F_0 contours in Mandarin seem to align with the syllable offset rather than the onset [Xu, 1994, 1997, forthcoming]. The findings of these studies suggest that such an alignment pattern may have to do with how sequences of lexical tones are implemented in production. As found in these studies, the implementation of each tone in a tone sequence always seems to start from the onset of the syllable and proceeds in such a way that the entire F_0 contour appropriate for that tone is continually approached until the end of the syllable, at which point *(and only at that point)* the implementation of the next tone start. Hence, the seemingly late alignment of the tonal contour to the syllable offset is probably in part due to this continual approximation of the tonal contour within a syllable. As mentioned earlier, however, since these studies examined only syllables with a simple CV structure, it is not yet known whether the same tone implementation pattern occurs in syllables with more complex structures. A similar tone implementation pattern occurring in more complicated syllable structures would provide stronger evidence that the entire syllable rather than anything else is the tone-bearing unit.

In Mandarin, in addition to the simple CV structure, there is also a CVN syllable structure, where N stands for the final nasal. Xu [1989] reported an interesting property in syllables with a CVN structure in Mandarin. That is, the duration ratio of the final nasal to the vowel (N / V) varies as a function of vowel height: the higher the vowel, the greater the value of N / V. This property makes it possible to compare F_0 contour alignment not only in syllables with or without a final nasal, but also in CVN syllables with different N-V ratios. The present study thus made use of this property and conducted two experiments to compare F_0 contour alignment in syllables with three potentially different N-V duration ratios – N-V ratio = 0 (i.e. when there is no final nasal), low N-V ratio (i.e. when the nuclear vowel is relatively low and, therefore, longer) and high N-V ratio (i.e. when the nuclear vowel is relatively high and short). Comparisons were also made of F_0 contours produced at three different speaking rates – normal, fast and slow. Experiment 1 investigated whether the F_0 contour of the entire rhyme or that of the vowel portion of a syllable alone is more consistent across different syllable structures. Experiment 2 examined F_0 contour alignment when final nasals are geminated with a following initial nasal or lateral. These experiments sought to answer the following questions: (a) Are the F_0 contours of the entire rhyme or that of the vowel portion of a syllable alone more consistent across different syllable structures? (b) When a final nasal is geminated with the following initial nasal or lateral, does the F_0 peak associated with the R tone occur in the later or earlier half of the nasal geminate? (c) As the duration of the final nasal increases, does the peak in the nasal geminate occur later or maintain a constant distance from the preceding vowel? (d) As the syllable duration increases, is the F_0 contour associated with the R tone stretched over the entire voiced portion of the syllable, or does the most appropriate contour shift later with the syllable offset? By addressing these questions, it is hoped that evidence will be found to indicate whether lexical tones in Mandarin are aligned with the syllable as a whole (without regard to its internal structure) and whether a contour tone such as R is inherently dynamic or composed of sequences of static pitch targets.

Experiment 1

This experiment tests whether the F_0 contour of the entire rhyme of a syllable or that of the vowel alone is more consistent across syllables with different N-V ratios and across different speaking rates.

Method

Stimuli. Table 1 lists all the target words used in the experiment. As shown in this table, the first syllable in each word has one of three possible N-V ratios: zero, low or high. The first syllable in each word has one of four different tones: H, R, L and F. These words were to be produced with the carrier sentence 'wǒshuō ... zhège cí' (I say the word ...). Three speaking rates – normal, fast and slow – were used.

Subjects. Four native speakers of Mandarin, two males (including the author) and two females, participated as subjects. Except for the author, all subjects were graduate students studying at Northwestern University. All subjects except the author had lived in Beijing, China, since childhood and spoke Beijing Mandarin. The author is a native speaker of Standard Chinese which closely resembles Beijing Mandarin phonetically.

Procedures. Recording was conducted in a sound-treated booth in the Speech Acoustics Laboratory at the Department of Communication Sciences and Disorders, Northwestern University. The speech signal was received by a condenser microphone connected to a Macintosh computer. The signal was digitized at 22 kHz in real time using the SoundEdit program (Macromedia Inc.) and stored on the computer's hard disk.

Table 1. Target words, carrier frame, speaking rates and total number of utterances in experiment 1

Expected N-V ratio	Words			
Zero	fāhuǒ	héshuǐ	hǎishàng	hàofèi
	fēifǎ	huáxuě	hǎohàn	hòuhuàn
Low	fāngfǎ	fánsuǒ	fǎnfù	fànshàng
	shānshuǐ	hánghǎi	fǎngxiào	huànsuàn
High	fēngshuǐ	hóngshuǐ	hěnhuài	fèngxiàn
	fēnxiǎng	xíngshǐ	shěnxùn	hùnfàn

The words are spelled in Pinyin. In this table and in subsequent tables and figures, tones are marked by diacritics placed on the nuclear vowel in each syllable. For example, in the syllable /ma/, the H, R, L and F tones are marked as fā, fá, fǎ and fà, respectively.

Carrier: wǒshuō ... zhège cí. I say the word ...

Speaking rate: normal, fast, slow.

Number of utterances: 3 (N-V ratios) × 4 (tones) × 2 (words) × 3 (speaking rates) × 5 (repetitions) = 360 (utterances).

The target words together with the carrier sentence were presented in Chinese on a Macintosh computer using a program written in Java. The subject was seated comfortably in the booth in front of a computer monitor. The microphone was placed by the side of the monitor, approximately 30 cm away from the subject's lips. At the start of the session, the subject clicked the 'start' button on the monitor screen with a mouse. A set of instructions was then displayed on the screen. After reading the instructions, the subject did a couple of practice trials before starting the real trials. In each trial, a sentence printed in Chinese appeared on the screen with the target word underscored, indicating that the underscored item should be stressed (in a Mandarin disyllabic word, both syllables are about equally stressed). The subject was instructed to read the target sentence aloud 3 times, first at a normal speaking rate, then at a fast rate and finally at a slow rate. The subject was also instructed not to pause in the middle of the sentence even at the slow speaking rate. In case a mistake was made, the subject was asked to repeat the same sentence at all three speaking rates. The sentences were presented in random order together with the sentences to be examined later in experiment 2, and the order of stimulus presentation was different for each subject.

F_0 Extraction. The digitized signals were transferred from the Macintosh computer to a Sparc 5 workstation. The signals were then converted to a format readable by programs in the ESPS/waves+ signal processing software package (Entropic Inc.). The individual target sentences were extracted and saved as separate ESPS signal files. The program *epochs* in the ESPS package was used to mark every pitch period in the target words. After that, the marked signals were manually edited in the ESPS *xwaves* program to correct spurious vocal pulse markings such as double marking or pitch period skipping. At the same time, segmentation labels were added to mark the boundaries between segments in the target word using the *xlabel* program.

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

Duration Analysis. The first set of analyses was to establish duration patterns that were relevant for later F_0 contour analysis. Three sets of duration measurements were taken from the segmented label files: (1) vowel duration, (2) nasal duration, (3) duration of the entire rhyme and (4) duration of the entire syllable. Table 2 displays, for each subject, mean rhyme duration and syllable duration of syllable 1 averaged by (a) speaking rate, (b) tone and (c) syllable structure. Probability values obtained in mixed repeated-measure ANOVAs are displayed at the bottom of the column for each subject. The duration values displayed in table 2a show that the subjects indeed followed the instructions and produced utterances with significantly different durations for the target syllables. The dura-

Table 2. Mean rhyme duration (left figure in each cell) and syllable duration (right figure in each cell) in milliseconds of syllable 1 from four subjects averaged by (a) speaking rate, (b) tone and (c) syllable structure

	XY	ZL	ZJ	ZM
a. Speaking Rate				
Normal	190, 305	172, 282	160, 259	161, 272
Faster	130, 188	103, 174	128, 201	124, 202
Slower	325, 491	233, 386	257, 398	206, 364
d.f.=2, 24	p<0.001, <0.001	p<0.001, <0.001	p<0.001, <0.001	p<0.001, <0.001
b. Tone				
H	214, 329	165, 278	182, 282	174, 281
R	231, 243	187, 301	196, 303	176, 301
L	214, 325	164, 271	174, 279	139, 256
F	199, 315	161, 275	175, 279	164, 279
d.f.=3, 12	p=0.002, =0.051	p=0.003, <0.001	p=0.048, =0.025	p<0.001, <0.001
c. Syllable Structure				
CV	190, 304	161, 274	165, 270	148, 269
CVN (with lower V)	234, 346	178, 284	194, 296	181, 290
CVN (with higher V)	220, 334	168, 284	186, 292	161, 278
d.f.=2, 12	p<0.001, <0.001	p=0.014, =0.100	p=0.002, =0.006	p<0.001, =0.005

Probability values obtained in a mixed repeated-measure ANOVA are displayed at the bottom of the column for each subject. The within factor is speaking rate. The between factors are tone and syllable structure. In this and the following tables, XY and ZL are male speakers, while ZJ and ZM are female speakers.

Table 3. Mean N-V ratio of syllable 1 from four subjects averaged by (a) speaking rate, (b) vowel height and (c) nasal place of articulation

	XY	ZL	ZJ	ZM
a. Speaking Rate				
Normal	1.12	1.57	0.77	0.67
Faster	0.79	1.40	0.44	0.39
Slower	1.31	1.49	0.56	0.80
d.f.=2, 24	p<0.001	p=0.700	p<0.001	p=0.002
b. Vowel Height				
Lower	0.65	0.80	0.38	0.34
Higher	1.50	2.18	0.80	0.89
d.f.=1, 12	p<0.001	p=0.003	p=0.004	p=0.002
c. Nasal Place				
Alveolar	0.84	1.00	0.47	0.45
Velar	1.38	2.11	0.75	0.84
d.f.=1, 12	p=0.007	p=0.023	p=0.073	p=0.042

Probability values obtained in a mixed repeated-measure ANOVA are displayed at the bottom of the column for each subject. The within factor is speaking rate. The between factors are vowel height and nasal place of articulation.

tion patterns for different tones shown in table 2b agree with the duration patterns reported previously [Xu , 1997]. In general, the duration of the R tone is the longest, and that of the H tone the second longest. The durations of the L and F tones are not consistently different from one another. Table 2c shows that, except for one subject (ZL), both rhyme duration and syllable duration differed significantly depending on the internal structure of the syllable in terms of whether it is CV or CVN or whether the nuclear vowel (in the CVC syllables) is higher or lower. In general, CVN syllables are longer than CV syllables, indicating that it takes longer to implement an extra segment in the syllable. The fact that syllables with lower nuclear vowels are longer reflects the well-known phenomenon of intrinsic duration [Lehiste, 1970].

Table 3 displays, for each subject, the mean N-V ratio of syllable 1 averaged by (a) speaking rate, (b) vowel height and (c) nasal place of articulation. Probability values obtained in a mixed repeated-measure ANOVA are displayed at the bottom of the column for each subject. Table 3a shows that, except for subject ZL, the N-V ratio varied significantly with change of speaking rate. The general trend seems to be that the N-V ratio (and also nasal duration) increases as speaking rate is reduced. Table 3b demonstrates that for all subjects, the N-V ratio is greater with higher nuclear vowels than with lower ones, confirming the observation of Xu [1989]. For all subjects, the N-V ratio increased more than twofold from lower vowels to higher ones. Table 3c shows that the N-V ratio is also mildly dependent on nasal place of articulation. For three subjects, the N-V ratio is higher when the nasal is velar than when it is alveolar. For subject ZJ, there is a significant interaction between vowel height and nasal place of articulation. Her N-V ratio for the velar is higher than for the alveolar when the vowel is higher, but the N-V ratio is not much different when the vowel is lower.

An important observation in table 3 is that there appear to be extensive differences among the four speakers in terms of their N-V ratios. In particular, the N-V ratio values are much higher for the two male speakers (XY and ZL) than for the two female speakers. Since there are only four subjects, however, it cannot be determined at this point whether this apparent sex difference is reliable. Regardless, it would be interesting to further examine whether the N-V ratio differences affect how F_0 contours align with syllables and segments within syllables.

F_0 Contour Analysis. To graphically examine F_0-syllable alignment, the smoothed F_0 curves were averaged across the 5 repetitions spoken at the same speaking rate using the following procedure. First, for each segment, a mean F_0 contour was computed by taking a fixed number of F_0 points at equal time intervals and averaging them across the 5 repetitions. Next, the mean duration of each segment was computed across the 5 repetitions. The mean F_0 curves of all segments were then plotted as a function of real time and aligned by different segment locations for visual inspection. Because the only thing averaged out when obtaining these plots was the variability in F_0 and duration across the 5 repetitions, everything that was consistently produced across the repetitions was fully preserved in these plots. Figure 1 displays, for all subjects, F_0 contours aligned by the offset of the nuclear vowel in syllable 1. Only half of all the F_0 plots are displayed in figure 1 due to space limitations and a high degree of similarity among the F_0 curves. To avoid any potential bias in data presentation, the words displayed in figure 1 were chosen alphabetically. They were all from the first row of each N-V ratio group displayed in table 1, which is alphabetically arranged.

Several observations can be made from the F_0 contour plots in figure 1. The first observation is that there seem to be individual differences in the overall F_0 height produced at different speaking rates, particularly in the case of the H tone (fig. 1a). Two of the subjects (XY and ZJ) produced higher F_0 values at faster speaking rates than at slower speaking rates, while the other two subjects produced similar F_0 heights across speaking rates. The second observation is that for most of the subjects, there is usually an F_0 drop around the onset of an F_0 contour, which is attributable to the F_0-raising effect of the initial voiceless fricative. The third and the most interesting observation is that, regardless of the syllable structure, the F_0 contour of the entire voiced portion of a syllable rather than just that of the vowel portion seems to be more consistent across different syllable structures. For all the four tones, the only real difference due to syllable structure variation is in how the same F_0 contour is distributed between the vowel and nasal portions of the rhyme, i.e. on both sides of the dashed vertical line in each plot. The longer the relative duration of the final nasal, the larger the portion of the F_0 contour occurring in the nasal segment.

For the H tone (fig. 1a), there are many individual variations in the F_0 contour shapes for the tone, but these shapes are consistent across different syllable structures. For the R tone (fig. 1b), there

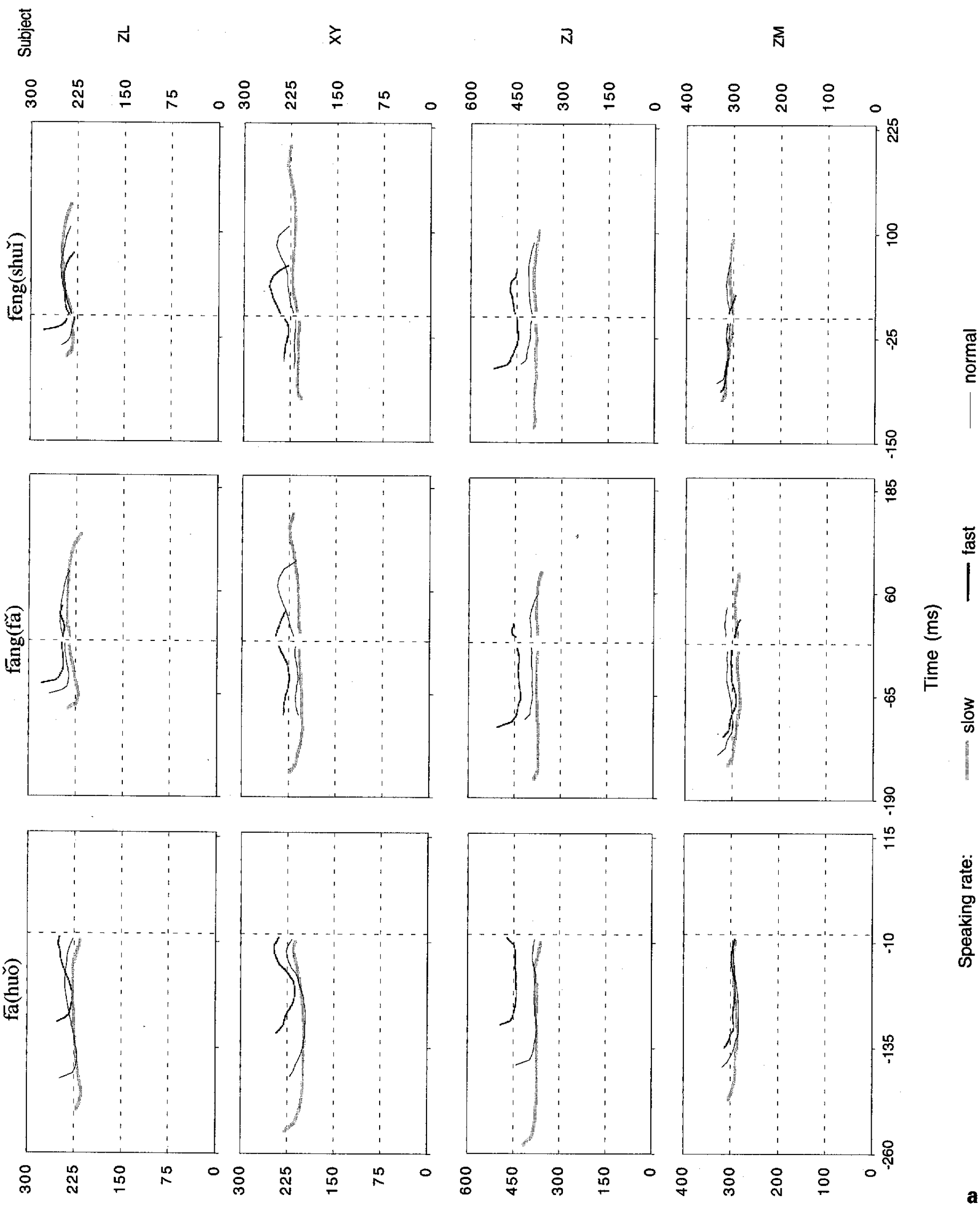
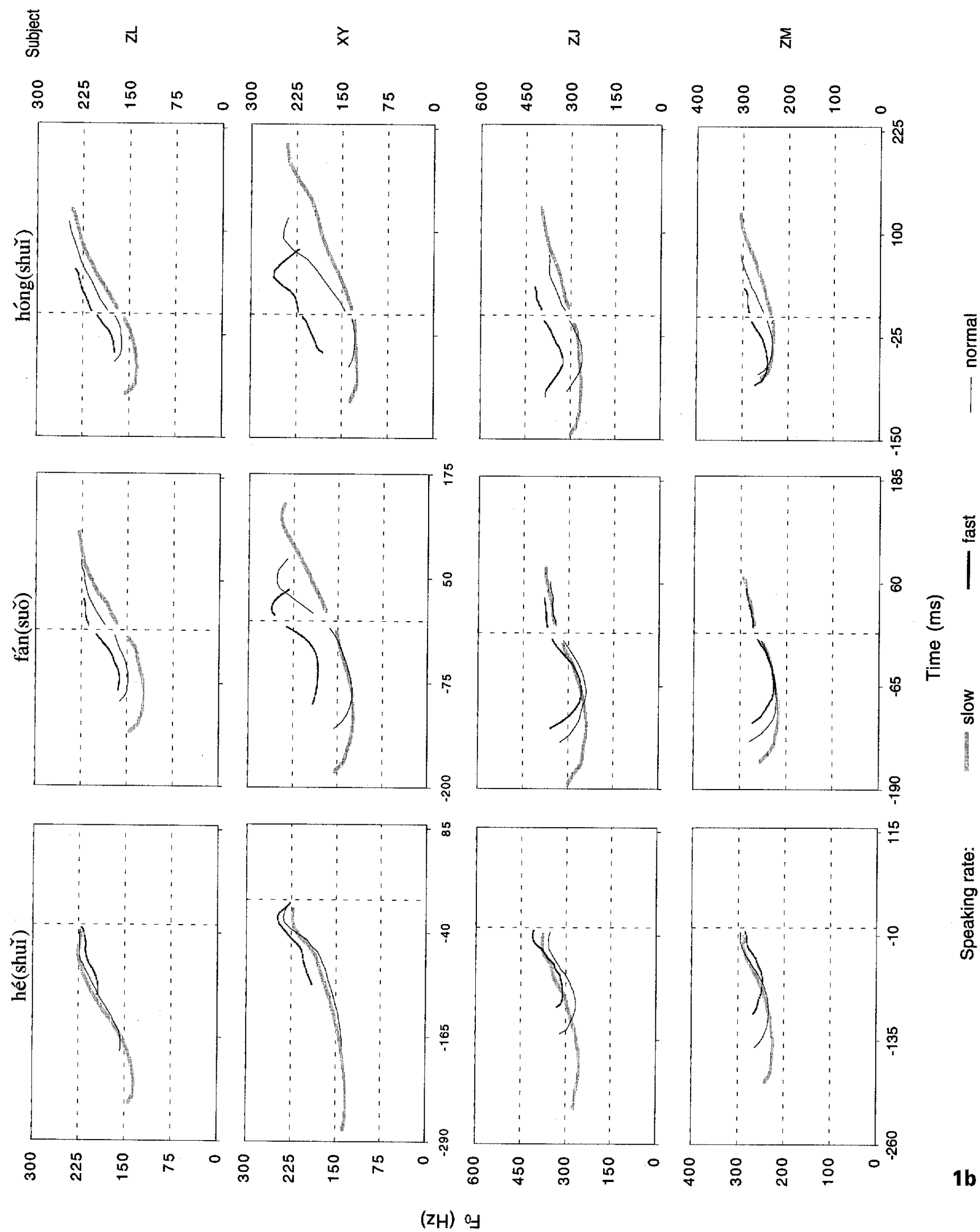
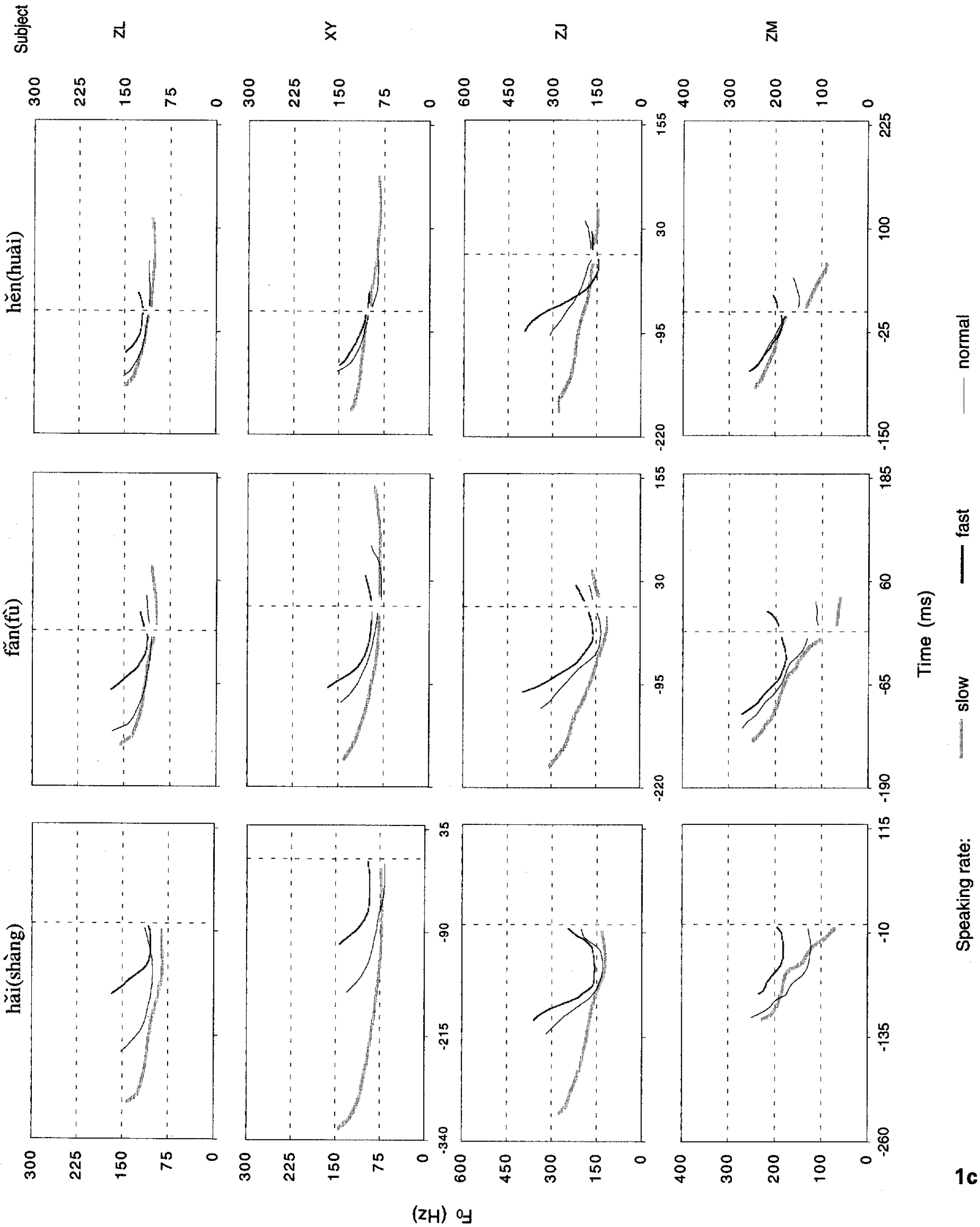
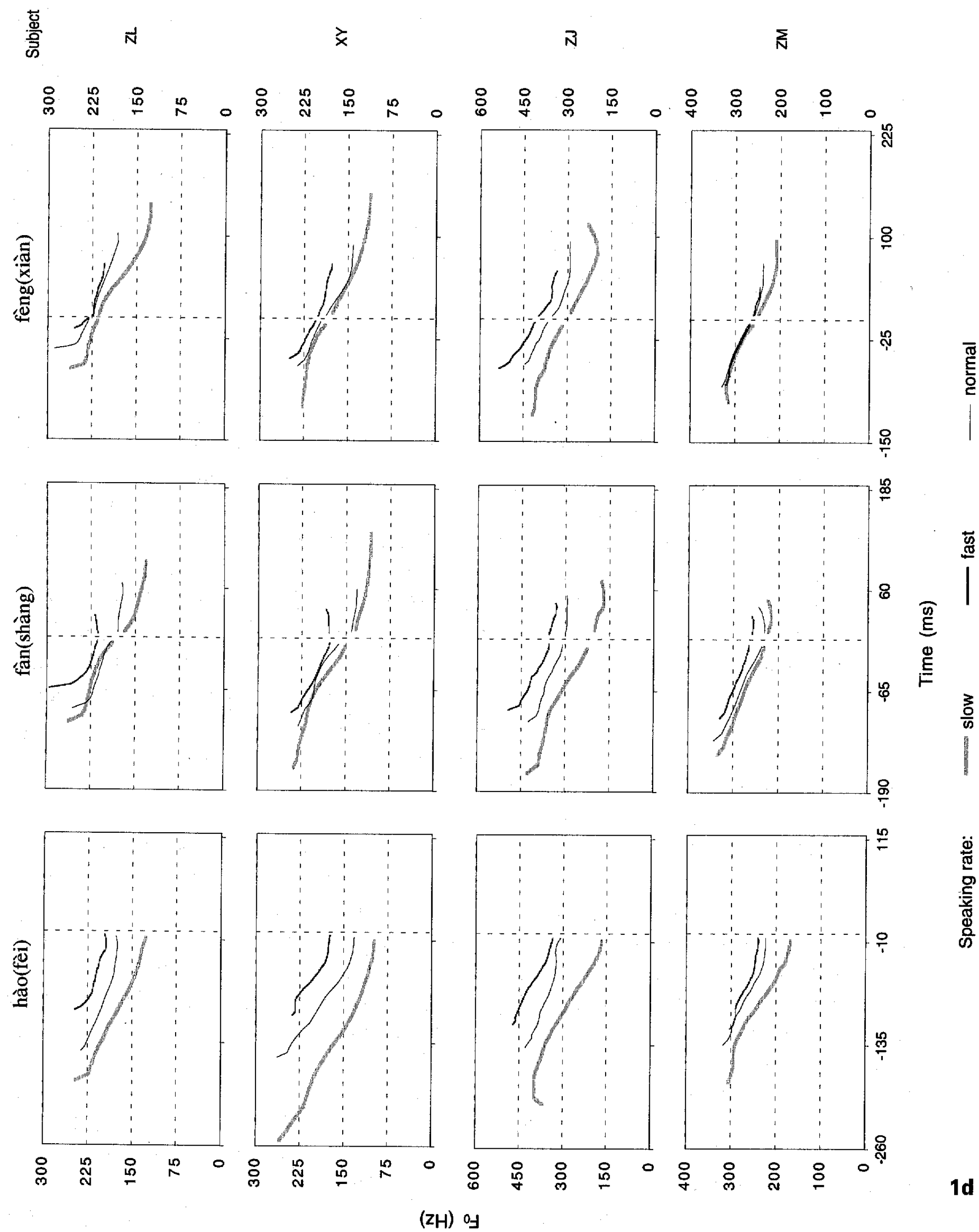


Fig. 1. Mean F_0 contours of syllable 1 in the disyllabic words of experiment 1 plotted as a function of time and aligned by the offset of the nuclear vowel (indicated by the dashed vertical line). In each graph, the ordinate is F_0 in hertz, and the abscissa is time in milliseconds (with 0 ms at the vowel offset). The four tones, H, R, L and F, are displayed in panels **a–d**, respectively. In this and the following figures, XY and ZL are the male subjects, while ZJ and ZM are the female subjects.


hé(shuǐ)
fán(suǒ)
hóng(shuǐ)
Subject
ZL
XY
ZJ
ZM
F0 (Hz)
Time (ms)
Speaking rate:
slow
fast
normal
1b


hǎi(shàng)
fǎn(fù)
hěn(huài)
Subject
ZL
XY
ZJ
ZM
F0 (Hz)
Time (ms)
Speaking rate:
slow
fast
normal
1c


Subject
ZL
XY
ZJ
ZM
hào(fèi)
fàn(shàng)
fèng(xiàn)
F0 (Hz)
Time (ms)
Speaking rate:
slow
fast
normal

1d

is usually an F_0 onset much lower than that of the H tone, which is followed by a sharp rise often all the way to the end of the syllable. When there is a final nasal, depending on its duration, part of or even the entire rising contour occurs during the final nasal. The L tone (fig. 1c) has an F_0 onset slightly lower than that of the R tone, followed by a sharp drop near the bottom of the F_0 range. The F_0 continues to fall gradually until the syllable offset. Again, this pattern remains consistent across syllables with and without a final nasal, and the relative duration of the nasal to that of the vowel affects only how this F_0 contour is distributed between the vowel and nasal portions of the syllable. The F tone (fig. 1d) usually has a high F_0 onset followed by a quick fall, but the F_0 never reaches the bottom as it does in the L tone (thus replicating similar findings made by Xu [1997] and Shih [1988]. The final portion of the F tone often has a reduced falling slope, but this is also consistent across syllables with and without a final nasal.

Discussion

The results of experiment 1 indicate that there is considerable individual variation in the duration of the final nasal relative to the vowel in a syllable as indicated by the N-V ratios. Despite the individual variation, however, the N-V ratio generally seems to be greater in syllables with a higher nuclear vowel than in those with a lower nuclear vowel, and in syllables with a final velar nasal than in those with a final alveolar nasal. Furthermore, regardless of the individual variations and variations related to vowel height and nasal place of articulation in terms of the N-V ratio, the F_0 contour for each tone seems to maintain a consistent alignment to the syllable that carries it. Variations in the N-V ratio seem to affect only how the entire F_0 contour is distributed across the vowel and nasal portions of the syllable: the longer the final nasal, the more of the final portion of the F_0 contour seems to be implemented during the final nasal.

Experiment 2

Experiment 1 demonstrated that when there is a final nasal in a syllable, the F_0 contour of a tone is distributed between the vowel and the nasal portions of the syllable. However, because the F_0 contour cannot continue during the initial fricative of syllable 2, the precise alignment of a tone with the host syllable as a function of syllable structure and speaking rate is difficult to assess. It would be preferable if F_0 contour alignment could be examined in syllables that are followed by a syllable starting with a vowel or glide, such as in 'yi' or 'ai'. Unfortunately, in Mandarin, when a final nasal is followed by a vowel- or glide-initiated syllable, the final nasal is usually realized as nasalization of the entire rhyme, and there is usually no nasal closure [Xu, 1986, 1989], making it difficult to accurately locate the syllable boundary. A situation in which nasal closure will not be affected while voicing continues across the syllable boundary is when a final nasal is followed by a syllable with an initial nasal (or/l/). In such cases, the two nasals (or the nasal and the /l/) will form a geminate (or pseudogeminate) with no clear syllable boundary in between [Xu, 1986, 1989]. It is therefore possible to compare the F_0 contour alignment in these disyllabic sequences to that of disyllabic sequences in which syllable 1 has no final nasal but syllable 2 has an initial nasal. Experiment 2 thus further investigated F_0 contour alignment by comparing the F_0 contours in words with geminated (or pseudogeminated) nasals and words with simple initial nasals produced at different speaking rates.

The first aim of experiment 2 was to determine the precise alignment of the most pertinent F_0 contour of a tone. To this end, a particular bitonal sequence, R-L, was

Table 4. Target words, carrier frame, speaking rates and total number of utterances in experiment 2

Expected N-V ratio	Words			
Zero	háomiǎo	hémǎ	húinuǎn	shúmǎi
	léiniǎo	luómǎ	niúmǎ	niúnǎi
Low	fánnǎo	hánlěng	huángmǎ	huángmǐ
	nánlǐng	nánměi	nánmiǎn	nánnǔ
High	hóngmǎ	hóngmǐ	shénniǎo	shénnǔ
	língmǐn	ménliǎn	mínglǎng	míngmǎ

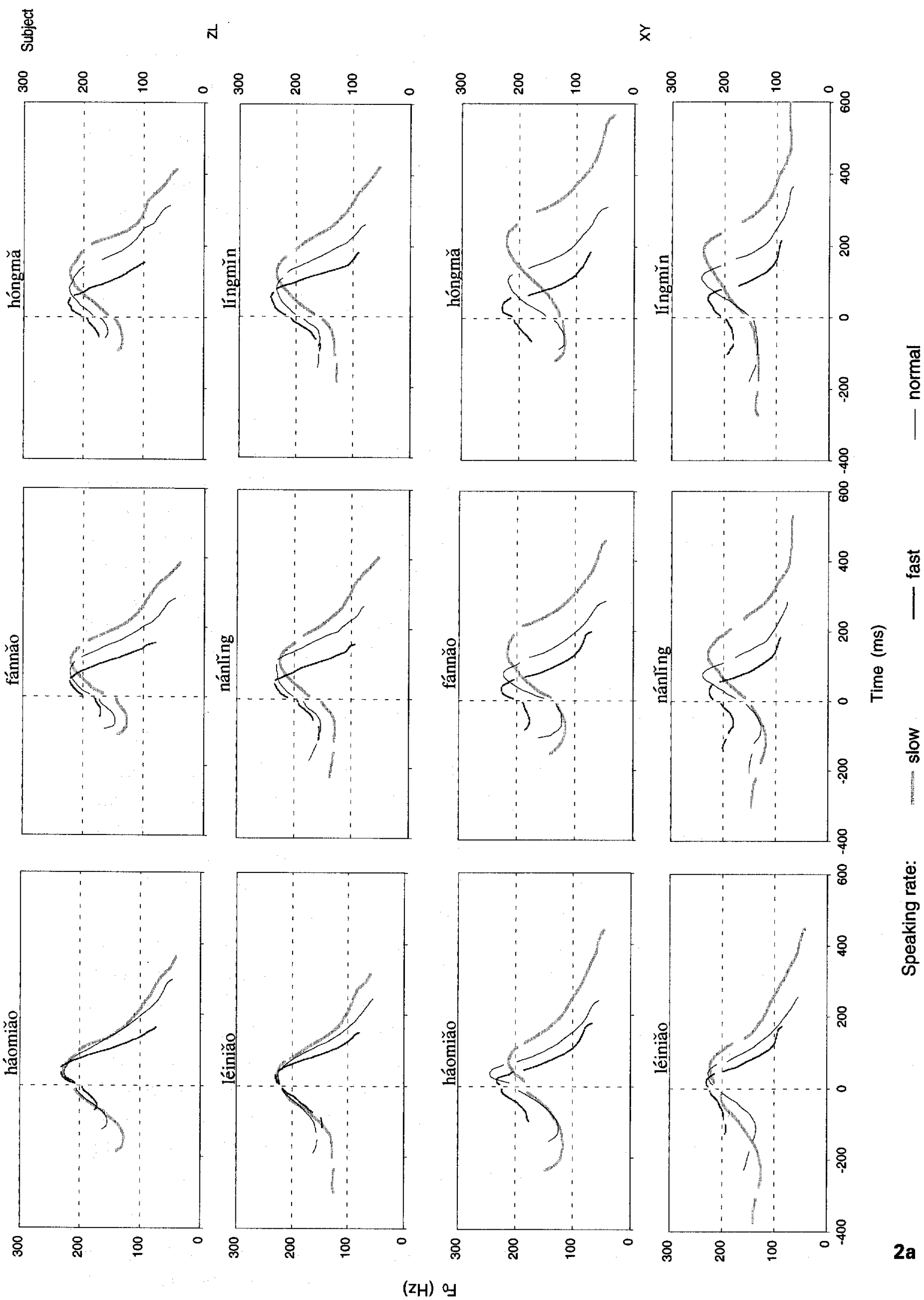
Carrier: wǒshuō … zhège cí. I say the word …
Speaking rates: normal, fast, slow.
Number of utterances: 3 (N-V ratios) × 8 (words) × 3 (speaking rates) × 5 (repetitions) = 360 (utterances).

examined. According to Xu [1997, forthcoming], when an R tone is followed by an L tone, the F_0 peak would occur on or just after the syllable offset when the R-tone-carrying syllable has no final nasal. If, as argued by Lin [1995], a tone is carried only by the vowel portion of a syllable, the F_0 contour most pertinent to the tone should occur within the vowel portion of the syllable regardless of how much of the syllable duration is occupied by the final nasal. In the case of an R-L sequence, the F_0 peak should occur on or just after the vowel offset, whether or not there is a final nasal and regardless of how long the duration of the final nasal is. If, however, the tonal contour is aligned to the entire syllable, the F_0 peak should occur on or just after the offset of the R-tone-carrying syllable, whether or not there is a final nasal in the syllable. In a disyllabic sequence in which the final nasal in syllable 1 is geminated with the initial consonant of syllable 2, the F_0 peak associated with the R tone should occur *inside* the geminated nasal segment, and its location should be in the later rather than the early half of the nasal geminate. Furthermore, the location of the peak should remain close to the vowel onset in syllable 2 (assuming the duration of the initial nasal or lateral of syllable 2 remains relatively constant) as the duration of the final nasal in syllable 1 varies, rather than remaining close to the vowel offset in syllable 1.

The second aim of experiment 2 was to determine whether the F_0 contour of the R tone in Mandarin is spread out evenly over the entire duration of the syllable as the syllable duration increases. To this end, alignment analyses were performed on portions of the F_0 contour before the final peak in the R tone. Three aspects of the earlier portions were examined, (a) the onset of the F_0 rise, (b) the slope of the rise and (c) the location of the maximum slope during the rise. If the F_0 contour is evenly spread as the syllable duration increases, the onset of the rise should remain close to the syllable onset, the slope of the rise should decrease, and the location of the maximum slope should at least remain in the middle of the syllable.

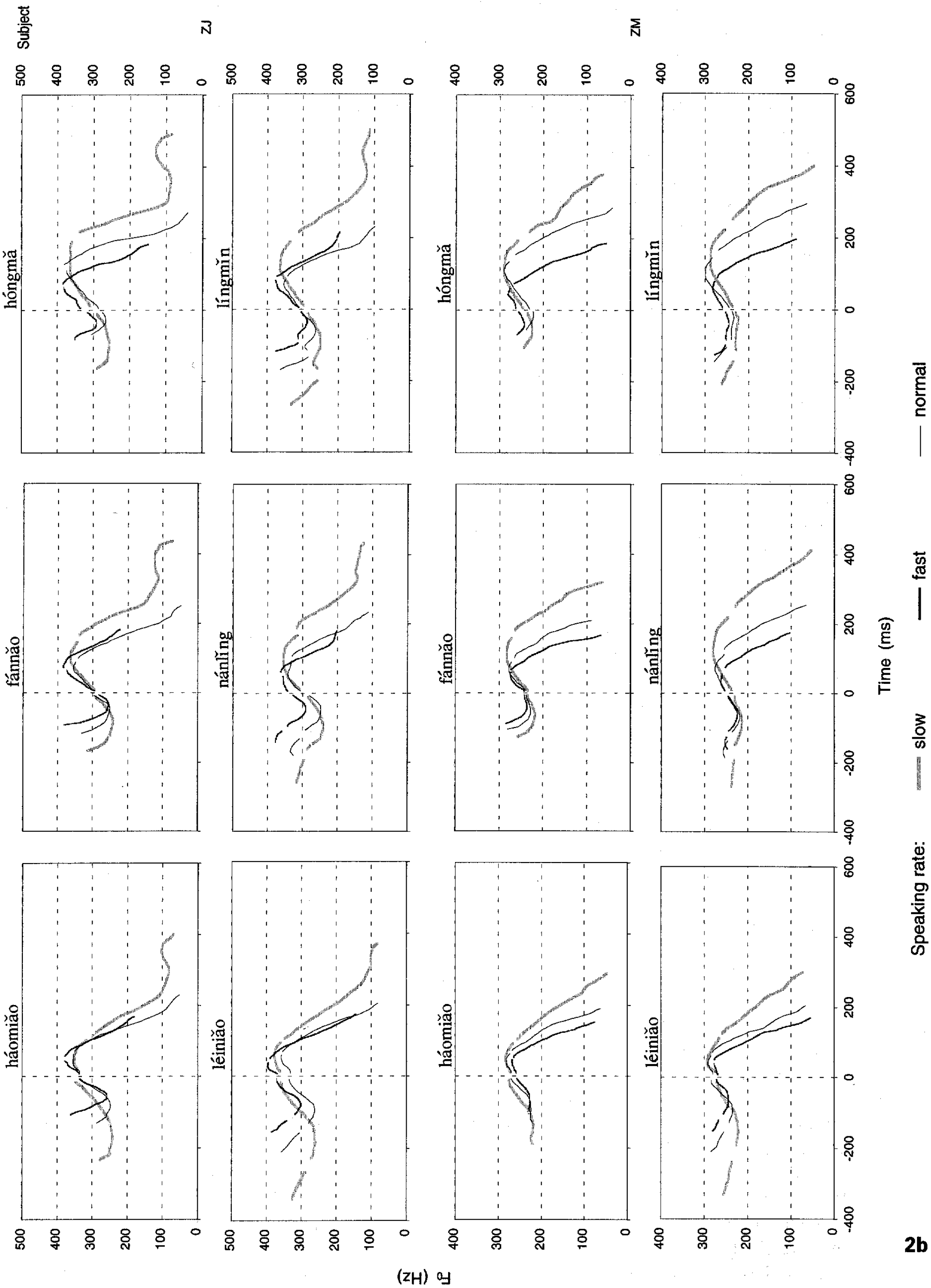
Methods

Stimuli. Table 4 lists all the target words used in experiment 2. As shown in table 4, the first syllable in each word has one of three possible N-V ratios: zero, low or high. These words all have the bitonal pattern of R-L. Half the words have initial sonorants (nasal or lateral) in syllable 1, white the other half have initial voiceless fricatives. As in experiment 1, these words were produced with the carrier sentence 'Wǒshuō … zhège cí' (I say the word …) at three speaking rates: normal, fast and slow.



2a

Fig. 2a, b. Mean F_0 contours of the disyllabic words in experiment 2 plotted as a function of time and aligned by the offset of the nuclear vowel in syllable 1. In each graph, the ordinate is F_0 in hertz, and the abscissa is time in milliseconds (with 0 ms at the vowel offset). The blank spaces in the F_0 curves correspond to segment boundaries.



Subject
ZJ
ZM
háomiăo
fánnăo
hóngmă
léiniăo
nánlĭng
lĭngmĭn
F0 (Hz)
Time (ms)
Speaking rate:
slow
fast
normal
2b

Subjects and Procedure. The same four speakers in experiment 1 served as subjects in experiment 2. The sentences for this experiment were recorded in random order together with the sentences for experiment 1. This way the potential monotony of the target words in this experiment was reduced because the tonal patterns of the words for experiment 1 were much more diverse.

F_0 Contour Extraction. An F_0 extraction procedure similar to that employed in experiment 1 was used to obtain smoothed F_0 curves. The smoothed F_0 curves were then averaged across the 5 repetitions using the same procedure as described for experiment 1. Also as in experiment 1, mean duration values were obtained for all segments and were used for graphically displaying the mean F_0 curves for visual inspection. Figure 2 displays, for all subjects, F_0 contours aligned by the offset of the nuclear vowel in syllable 1. Due to space limitations and a high degree of similarity among the F_0 curves, only a quarter of the obtained F_0 plots are displayed in figure 2. To avoid any potential bias in data presentation, the words displayed in figure 2 were chosen alphabetically. They were all from the first column of table 4, which is arranged alphabetically.

Visual Inspection. Several observations can be made by visually inspecting the F_0 contours in figure 2. First, for each subject, the F_0 curves of different words are very similar to one another, regardless of the difference in syllable structure and regardless of whether the initial consonant is a sonorant or a voiceless fricative. Second, as was observed in experiment 1, as the *relative* duration of the vowel in syllable 1 becomes shorter due to an increase in vowel height or due to a slower speaking rate, greater portions of the rising contour in syllable 1 occur in the geminated nasal (or nasal-lateral) segments, rather than the entire rising contour being increasingly squeezed into the relatively shortened vowels. Third, the F_0 peak generally falls on the nasal or the geminated nasal segment. However, for two of the subjects (ZL and ZJ), the peak occasionally occurs at the very beginning of the vowel in syllable 2 at the fast speaking rate. Fourth, as word duration increases with slower speaking rates, the most dynamic portion of the F_0 contour in syllable 1 seems to shift rightward, rather than being fully stretched out over the entire voiced portion of the syllable.

To examine the F_0 contour alignment more quantitatively, several sets of regression analyses were performed. The purpose of these analyses was to test (a) whether the F_0 peak associated with the R tone in syllable 1 aligned more closely to the boundary between syllables 1 and 2 or to the V-N boundary within syllable 1 and (b) whether the portion of the F_0 contour most appropriate for the R tone moved with the syllable offset or was spread out evenly as syllable duration increased.

Peak Alignment. To test whether the F_0 peak associated with the R tone in syllable 1 aligned more closely to the boundary between syllables 1 and 2 or to the V-N boundary within syllable 1, the location of the offset of syllable 1 needed to be determined. Although it is difficult to accurately locate the syllable offset due to the nasal gemination at the syllable boundary, the S-V boundary in syllable 2 may be used as a reference point (where S stands for the initial sonorant in syllable 2), assuming that the duration of the initial nasal does not vary systematically with any factor under scrutiny other than speaking rate. Two regression analyses were therefore conducted. The first used the distance between the onset of syllable 1 and the S-V junction in syllable 2 as the regressor, while the second used the distance between the onset of syllable 1 and the V-N junction in syllable 1 as the regressor. Both analyses used the location of the F_0 peak relative to the onset of syllable 1 as the dependent variable.

Figure 3 displays the regression plots for the first set of analyses. In this figure, the distance between the onset of syllable 1 and the F_0 peak is plotted against the distance between the onset of syllable 1 and the S-V boundary in syllable 2 for each subject. As can be seen, for all subjects, as the speaking rate decreases, the distance between the onset of syllable 1 and the S-V boundary in syllable 2 increases, indicating that the subjects indeed produced the target words with a wide range of duration values at different speaking rates. Also shown in figure 3 are the Pearson r^2 values. These r^2 values are very high: all greater than 0.94.

Since both sets of location values are measured from the onset of syllable 1, the distribution of the location of the F_0 peaks as a function of the location of the S-V boundary in syllable 2 indicates how closely the two are aligned with each other as overall duration changes with speaking rate. As indicated by the regression plots in figure 3, as the location of the S-V boundary in syllable 2 moves rightward as speaking rate slows down, the location of the F_0 peak also moves rightward, and does so consistently, as indicated by the high r^2 values.

The high r^2 values do not mean, however, that the F_0 peaks always stay near the S-V boundary in syllable 2. If they did, the slope of the regression lines would have been close to 1, and the regres-

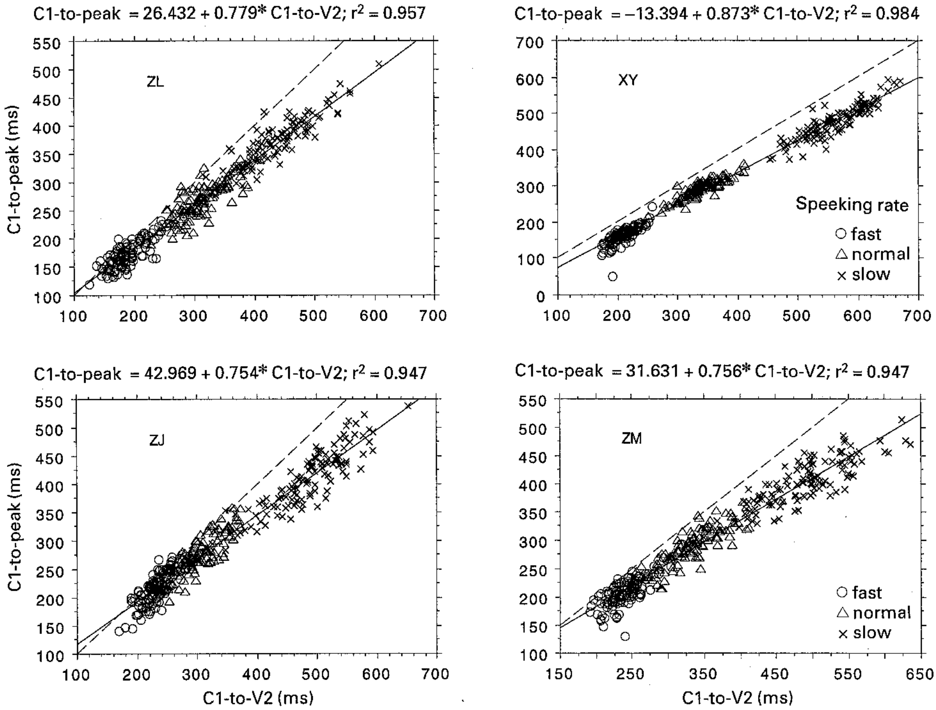


Fig. 3. Location of F_0 peaks relative to the onset of syllable 1 (y-axes) plotted against the distance between the onset of syllable 1 and the S-V boundary in syllable 2 (x-axes) for each subject. The plotting symbols represent different N-V ratios: circle for the fast rate, triangle for the normal rate and cross for the slow rate. The dashed line has a slope value of 1 and an intercept of 0.

sion line would have coincided with the dashed line in each graph. These dashed lines all have a slope value of 1 and an intercept value of 0. As indicated by the proximity of the regression line and the dashed line near the lower left corner of each graph, at the fast speaking rate, the F_0 peaks indeed occur very close to the S-V boundary in syllable 2. For subject ZL, a number of points fall on the upper side of the dashed line, indicating that some F_0 peaks actually occur *after* the S-V boundary inside syllable 2, corroborating what was observed in the plots of his F_0 contours in figure 2. As indicated by the values of the regression coefficients, which fall between 0.75 and 0.80 for all subjects, as speaking rate slows down, the F_0 peak occurs increasingly earlier than the S-V boundary in syllable 2. This trend can also be seen clearly in the F_0 plot in figure 2: as speaking rate decreases, the F_0 contour starts to fall increasingly earlier before the S-V junction in syllable 2.

The second set of regression analyses examined the relationship between the F_0 peak location and the V-N boundary inside syllable 1. Figure 4 displays the plots of these regressions for all subjects. In each graph, the locations of the F_0 peaks as measured from the onset of syllable 1 are plotted against the location of the V-N boundary in syllable 1, which is also measured from the syllable onset. For each subject, three separate regression lines are plotted, each corresponding to an N-V ratio group. The equations for these regression lines are displayed above each graph, and the Pearson r^2 values are displayed alongside the regression equations.

Unlike the almost linear clusters seen in figure 3, the points in figure 4 fan out as the distance between the V-N boundary and the onset of syllable 1 increases. As indicated by the plotting symbols and the regression lines, these points diverge into three groups according to the N-V ratios. The alignment pattern of each group is revealed by the coefficients of the regression equations. When the N-V

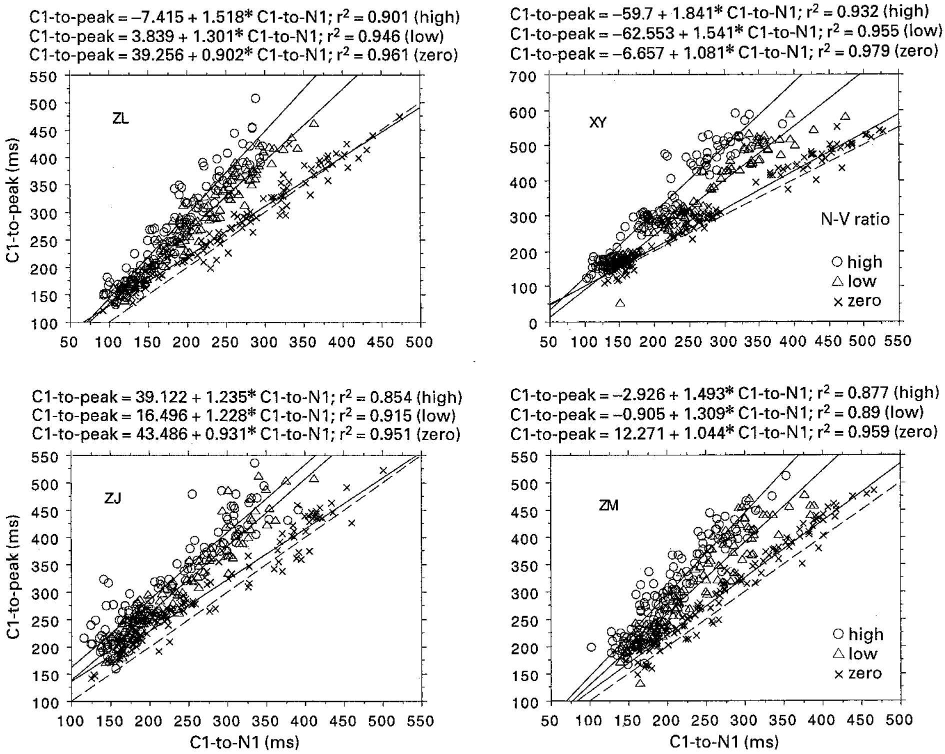


Fig. 4. Location of F_0 peaks relative to the onset of syllable 1 (y-axes) plotted against the distance between the onset of syllable 1 and the V-N boundary in syllable 1 (x-axes) for each subject. The plotting symbols represent different N-V-ratios. The dashed line has a slope value of 1 and an intercept of 0.

ratio is zero (i.e. when the syllable structure is CV), the coefficients are close to 1 for all subjects, indicating that F_0 peaks in these syllables stay very close to the V-N boundary, because the distance between the syllable onset and the V-N boundary in these words is equivalent to the duration of syllable 1. When the N-V ratio is greater than zero (i.e. when the syllable structure is CVN), the regression coefficients are much higher than 1 for all subjects and are higher for syllables with higher N-V ratios than those with lower N-V ratios. These coefficient values indicate that as the combined duration of the initial consonant and the vowel in syllable 1 increases, the F_0 peaks occur increasingly farther ahead of the V-N boundary in syllable 1, and the higher the N-V ratio, the faster the increase. Combined with the results of the first set of regression analyses, this seems to further indicate that the V-N boundary inside syllable 1 is a much less direct reference point for the location of the F_0 peaks than the S-V boundary in syllable 2.

Contour Alignment. To answer the question of whether the portion of the F_0 contour most appropriate for the R tone moves with the syllable offset or is spread out evenly as syllable duration increases, alignment analysis on portions of the F_0 contour before the final peak was needed. Three aspects of the earlier portions can be examined: (a) the onset of the F_0 rise, (b) the slope of the rise and (c) the location of the maximum slope during the rise. As may be seen in figure 2, it may not always be easy to take these measurements accurately. In particular, the location of the onset of the rise may be difficult to determine. As a first approximation, the onset of the rise may be defined as the F_0 minimum right before the rise. However, in some cases (in particular for subject ZL), as can be seen in

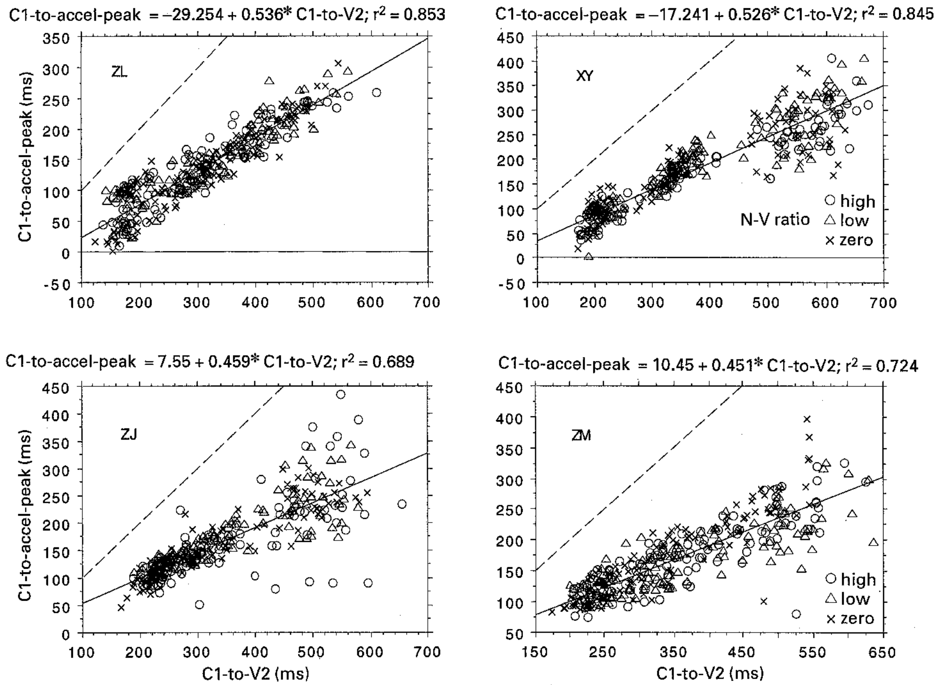


Fig. 5. Location of maximum acceleration at the onset of F_0 rise relative to the onset of syllable 1 (y-axes) plotted against the distance between the onset of syllable 1 and the S-V boundary in syllable 2 (x-axes) for each subject. The plotting symbols represent different N-V ratios. The dashed line has a slope value of 1 and an intercept of 0; accel = acceleration.

figure 2, the portion of the contour before the apparent rise is virtually flat and sometimes even rises slightly. In such cases, the F_0 minima seem too far away from where the real rise is. The real onset of the rise seems to be rather at the moment when the F_0 contour takes a sharp turn upward, i.e. when the F_0 movement has the greatest acceleration. Mathematically, the moment of greatest acceleration corresponds to the local maximum in the second derivative of the F_0 curve. As for the location and the value of the maximum slope in the F_0 rise, they correspond mathematically to the location and value of the maximum velocity, i.e. the first derivative of the F_0 curve. A computer program was thus written to compute the first and second derivatives of each F_0 curve and to locate the moments of maximum velocity and acceleration in the derivatives. To reduce noise in the derivatives, the F_0 curves (which had already been trimmed of sharp spikes and edges using the trimming algorithm [Xu, forthcoming]) were further smoothed by a triangular window function before the derivatives were taken.

The three measurements (i.e. the location of the maximum acceleration and the location of the maximum velocity, both relative to the onset of syllable 1, and the value of the maximum velocity) were then regressed to the distance between the onset of syllable 1 and the S-V boundary in syllable 2, which roughly represents the duration of syllable 1, as discussed earlier. Figures 5–7 display these regression plots for all subjects. The first two regressions, as shown in figures 5 and 6, respectively, reveal the location of the onset of the rise and the point of maximum velocity relative to the onset of syllable 1 as the duration of syllable 1 varies. The third regression (fig. 7) reveals how the maximum slope of the rise varies with the syllable duration.

In figure 5, the slopes of the regression lines are all around 0.5 (with those of the male speakers XY and ZL slightly steeper than those of the female speakers), indicating that the onset of the F_0 rise maintains about equal distance from the onset of syllable 1 and the S-V boundary in syllable 2. Since

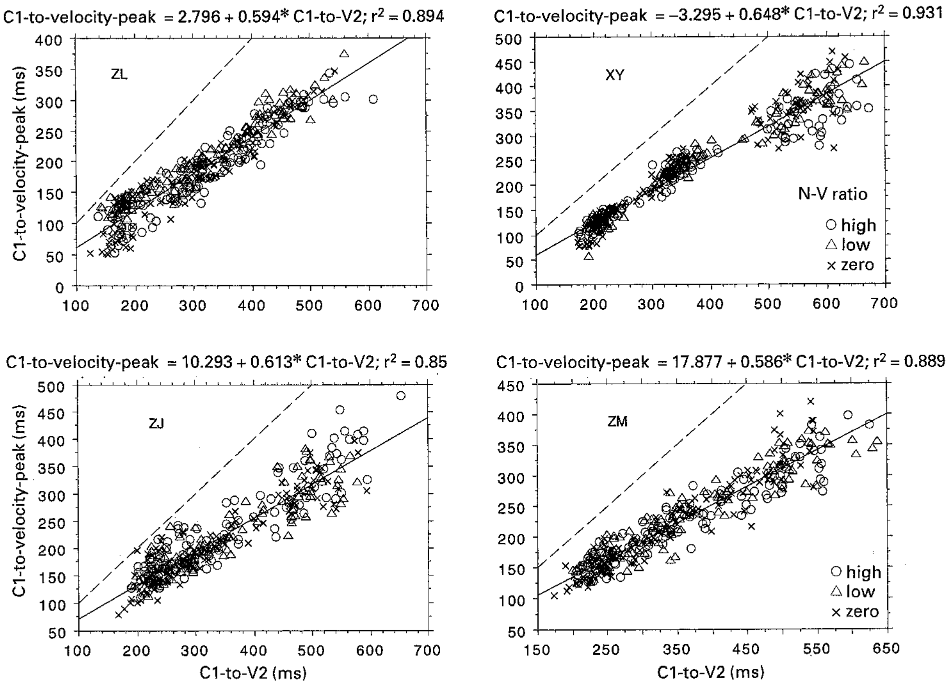


Fig. 6. Location of peak velocity of F_0 rise relative to the onset of syllable 1 (y-axes) plotted against the distance between the onset of syllable 1 and the S-V boundary in syllable 2 (x-axes) for each subject. The plotting symbols represent different N-V ratios. The dashed line has a slope value of 1 and an intercept of 0.

C1-to-V2 (the distance between the onst of syllable 1 and the S-V boundary in syllable 2) consists of both the entire duration of syllable 1 and the duration of the initial nasal of syllable 2, the onset of the F_0 rise probably actually moved slightly more in synchrony with the offset than the onset of syllable 1. To see whether this may have been the case, estimated durations of syllable 1 were computed. First, for each subject, a simple linear prediction equation was obtained for the duration of the initial sonorant in syllable 2 in words in which syllable 1 had no final nasal, using C1-to-V2 as the predictor. Then, the estimated duration of syllable 1 in each token was computed by subtracting the duration of the initial sonorant (predicted by the equation) from C1-to-V2. A new set of regression equations was then obtained using the estimated syllable 1 duration as the regressor and the location of the rise onset relative to the onset of syllable 1 as the dependent variable. These regression equuations are listed in table 5.

As can be seen in table 5, the coefficients of the regression equations are all above 0.5, ranging from 0.55 to 0.65. This indicates that if the actual duration of syllable 1 was as predicted, the onset of the F_0 rise would have moved more in synchrony with the offset than the onset of syllable 1. To see this more clearly, the locations of the rise onset at the estimated syllable duration of 150 ms were computed using these equations. When the estimated duration of syllable 1 is 150 ms, the predicted locations of the F_0 peaks relative to the onset of syllable 1 are 75, 85, 97 and 74 ms for ZL, XY, ZJ and ZM, respectively; and at the estimated syllable duration of 500 ms they are 289, 312, 290 and 272 ms, respectively. In other words, for all subjects but one (ZJ), as the estimated syllable duration increases, the onset of the rise moves further into the later portion of syllable 1.

In figure 6, the slopes of the regression lines range from 0.59 to 0.65 for all the subjects, indicating that the location of the velocity peak also moves more in synchrony with the offset than with

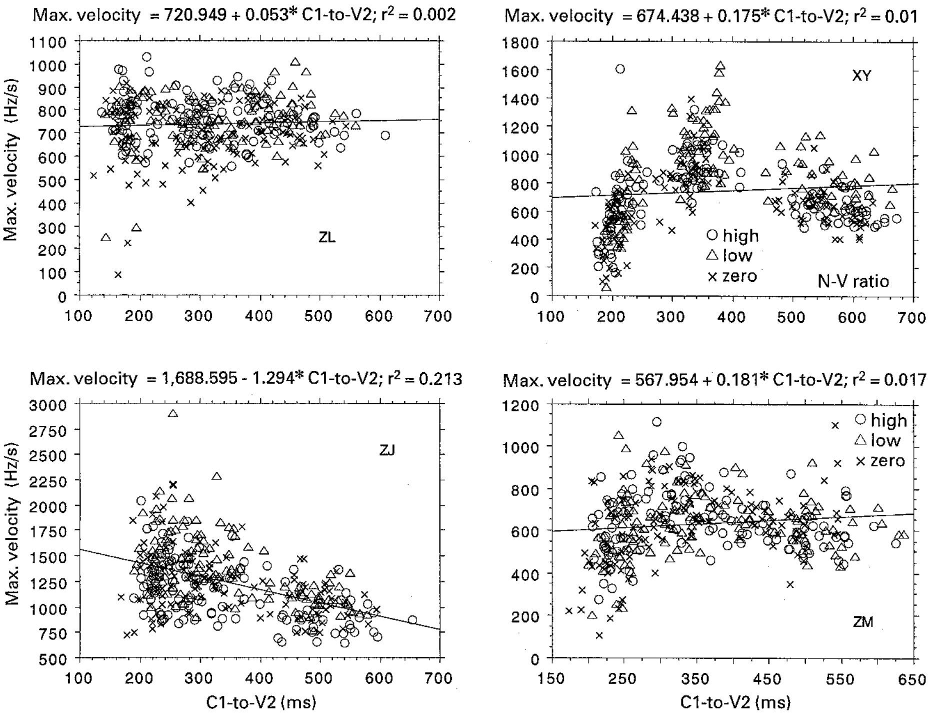


Fig. 7. Maximum rising velocity plotted against the distance between the onset of syllable 1 and the S-V boundary in syllable 2 (x-axes) for each subject. The plotting symbols represent different N-V ratios.

Table 5. Simple linear regression equations for the location of F_0 rise onset as a function of estimated duration of syllable 1

Subject	Linear regression equation	r^2
ZL	rise-onset = −16.256 + 0.611 × estimated syllable 1 duration	0.853
XY	rise-onset = −11.793 + 0.647 × estimated syllable 1 duration	0.844
ZJ	rise-onset = 14.39 + 0.551 × estimated syllable 1 duration	0.689
ZM	rise-onset = −10.504 + 0.565 × estimated syllable 1 duration	0.722

See text for the computation for the estimated syllable 1 duration.

the onset of syllable 1. In figure 7, the slopes of the regression lines are close to 0 for all subjects except ZJ, for whom the slope is −1.3. Since the units of the two axes are different (Hz vs. ms), the more meaningful slope value is the standard coefficient. For subject ZJ, the standard coefficient is −0.46. For the other three subjects, it is between 0.13 and 0.05. Therefore, except for ZJ, the slope of the rise in the F_0 contour bears little relationship with the duration of the syllable.

The regression plots in figures 5–7 also used different symbols to indicate the three N-V ratios. As can be seen, once again, no relationship is evident between the N-V ratio and any of the three measurements displayed in the figure.

Discussion

The results of the F_0 peak alignment and contour alignment analyses together with visual inspection of the F_0 contours reveal some interesting patterns of alignment between the F_0 contour and the syllable. Regardless of the syllable structure, the F_0 peak associated with the R tone always occurs close to the offset of the host syllable; the onset of the F_0 rise always occurs around the center of the host syllable (and possibly moves more in synchrony with the syllable offset than the onset); the peak velocity occurs closer to the syllable offset than the onset, and the maximum velocity of the F_0 rise does not vary consistently with either syllable duration or syllable structure. These results indicate that the internal structure of the syllable bears no direct relation to the F_0 contour alignment. They also indicate that the F_0 contour of the R tone in Mandarin does not spread out evenly as the duration of the host syllable increases but rather shifts as a whole toward the later portion of the syllable.

General Discussion

The results of the two experiments in the present study have implications related to several lines of research mentioned in the Introduction. First, concerning how lexical tones are aligned to the segmental units in Mandarin, the present results, together with findings reported previously [Xu, 1997, forthcoming], suggest that the syllable, rather than any other unit, is the reference domain for proper tonal alignment. Howie [1974] has argued that because there is much F_0 perturbation in the early portion of a syllable due to the initial consonant, the domain of a tone is limited to the rhyme of the syllable. As found by Xu, [1997, forthcoming], however, because the F_0 contour during the entire interval of a syllable is continuously approaching the most ideal contour for the tone assigned to the syllable, it is natural that the most appropriate F_0 contour occurs in the later rather than in the early portion of the syllable. In addition, because the early portion of the F_0 contour is subject to perturbation due to both the initial consonant and the carryover influence of the preceding tone, it is also natural that the most conspicuous perturbations of the F_0 contour occur in the early portion of the syllable. The finding that immediately after the offset of a syllable the F_0 contour begins to move toward the first target of the next tone [Xu, 1997, forthcoming] indicates that the implementation of a tone starts from the onset of its host syllable rather than from the onset of the rhyme. The results of the present study demonstrate that this tonal alignment pattern is not affected by the internal structure of the syllable in any direct way, further confirming the hypothesis that the syllable is the only relevant unit for the alignment of lexical tones in Mandarin.

Second, the results of the peak alignment analyses in the present study basically replicate the findings of Xu [forthcoming]. That is, the F_0 peak in the R tone always occurs near the offset of the syllable carrying the tone. The results of experiment 2 demonstrate that this is true even when the syllable offset occurs in the midst of a nasal or nasal-lateral geminate: the F_0 peak occurs right before the sonorant-vowel boundary in syllable 2 whether or not the initial sonorant in syllable 2 is geminated with the final nasal of the preceding syllable. This finding may have useful implications for speech processing – speech recognition in particular. In Mandarin, there are many cases where the syllable boundary is blurred by continuous voicing plus lack of abrupt spectral change. As found by Xu [1986], the final nasal of a syllable usually becomes mere nasalization of the preceding nuclear vowel when the following syllable starts with a vowel or a glide, and the two syllables usually appear to blend into each other seamlessly at the boundary because there is no abrupt spectral change, e.g. between 'mián' and 'ǎo' in the word 'miánǎo' (cotton-padded jacket). In such cases, the F_0 peak may be used as an indicator for the location of the actual syllable boundary.

Third, regarding the question of whether phonetically dynamic F_0 contours such as rising and falling are composed of simple underlying pitch targets such as H and L or are primary units themselves, the present results provide some interesting evidence. The regression analyses in experiment 2 show that: (a) the most dynamic portion of the rising contour moves increasingly into the later portion of the syllable as the duration of the syllable increases; (b) the onset of the rise stays roughly in the center (and possibly moves more in synchrony with the syllable offset than onset) of the syllable as its duration increases; (c) the maximum velocity of the rise does not change systematically with variations in syllable duration. These results demonstrate that the low and high points in the rising contour are not distributed evenly over the entire syllable. Instead, the rising contour as a whole shifts more into the later portion of the syllable without systematic changes in the slope of the rise as the syllable duration increases. This can be interpreted as evidence that a coherent rising contour is being maintained. In addition, the fact that there are extensive random variations in the slope of the rise at any given speaking rate suggests that speakers may be trying to produce a coherent dynamic contour rather than any particular rate of F_0 change. As for whether a contour tone such as R should still be considered a sequence of L and H, it is conceivable that one could perhaps argue that an LH sequence can be moved as a unit. However, such a unit would still have two internal targets by definition. To produce a sequence of two targets, the speaker would take full advantage of the excessive time available to the tone-carrying syllable when the speaking rate is reduced and make each target more individually salient. The fact that speakers instead move the whole contour further into the later portion of the syllable as duration increases suggests that there are probably not two separate targets but a single dynamic target.

Finally, while the findings of the present study are about how tones are implemented in production, they may also have certain implications in understanding how tones are perceived by listeners. Assuming that there is a direct link between perception and production [Liberman and Mattingly, 1985], what is observed or proposed about production should also be directly relevant to perception and vice versa. As mentioned in the Introduction, the tone perception model proposed by House [1985, 1990] would predict that a dynamic F_0 contour occurring across a V-N boundary should be heard as two static tones rather than as a single dynamic tone. This means that to produce a perceptually acceptable R tone in a syllable with a final nasal, speakers would have to implement the most dynamic portion of the rising contour inside the vowel rather than over the entire rhyme. As the results of both experiments in the present study demonstrate, however, this is clearly not the case in Mandarin. Instead, the most critical portion of an F_0 contour is found to occur across the V-N boundary in the rhyme, despite the rapid spectral change that takes place during the vowel-nasal shift, indicating that the syllable boundaries rather than regions of rapid spectral change are the most likely reference points used by speakers. While it remains to be seen whether this is also the case in other languages, at least one study [von Santen and Hirschberg, 1994] has found that for the ending F_0 contour of a phrase in English, the F_0 value at the end of a pre-sonorant vowel is much more variable than at the end of the last sonorant. This indicates that English speakers, too, probably implement the final pitch target by the end of the last sonorant rather than by the end of the pre-sonorant vowel.

Conclusion

In two experiments, the shape and alignment of F_0 contours in syllables with and without a final nasal were examined at normal, fast and slow speaking rates through graphic comparison and contour alignment analysis. Results indicate that the internal structure of a syllable bears no direct relation to the alignment of the tonal contours with their host syllables. Regardless of whether the syllable contains a final nasal or not, the shape and alignment of the F_0 contours are most consistent across the entire syllable rather than just within the vowel portion of the syllable. Furthermore, as the duration of a syllable increases, the F_0 contour most appropriate for the tone associated with the syllable does not spread out evenly over the entire syllable duration but rather moves increasingly into the later half of the syllable. These results were interpreted as indicating that the syllable is the unit of reference for tonal alignment (or domain of tone implementation) in Mandarin, and that tone contours such as rising and falling are probably implemented as integral dynamic targets rather than as sequences of static targets.

Acknowledgment

I would like to thank Bruce Smith for comments on an earlier version of this paper. My thanks also go to Arthur S. Abramson and an anonymous reviewer for their comments. This study was partly supported by the Volwiler Research Fund.

References

Abramson, A.S.: The phonetic plausibility of the segmentation of tones in Thai phonology. Proc. 12th Int. Congr. Ling., pp. 760–763 (Vienna 1978).

Anderson, S.R.: Tone features; in Fromkin, Tone: a linguistic survey, pp. 133–175 (Academic Press, New York 1978).

Bolinger, D.L.: Intonation: levels versus configuration. Word *7:* 199–210 (1951).

Clark, M.: A dynamic treatment of tone, with special attention to the tonal system of Igbo (University of Massachusetts, Amherst 1978).

Duanmu, S.: Against contour tone units. Linguistic Inquiry *25:* 555 (1994).

Gandour, J.: On the representation of tone in Siamese. UCLA Working Papers Phonetics *27:* 118–146 (1974).

House, D.: Implications of rapid spectral changes on the categorization of tonal patterns in speech perception. Working Papers, Department of Linguistics and Phonetics, Lund University *28:* 69–89 (1985).

House, D.: Tonal perception in speech (Lund University Press, Lund 1990).

Howie, J.M.: On the domain of tone in Mandarin. Phonetica *30:* 129–148 (1974).

Ladd, D.R.: The structure of intonational meaning; PhD diss. Cornell University (1978).

Leben, W.R.: Suprasegmental phonology; PhD diss. Massachusetts Institute of Technology (1973).

Lehiste, I.: Suprasegmentals (MIT Press, Cambridge 1970).

Liberman, A.M.; Mattingly, I.G.: The motor theory of speech perception revised. Cognition *21:* 1–36 (1985).

Liberman, M.; Pierrehumbert, J.: Intonational invariance under changes in pitch range and length; in Aronoff, Oehrle, Language sound structure, pp. 157–233 (MIT Press, Cambridge 1984).

Lin, M.: A perceptual study on the domain of tones in standard Chinese. Chinese J. Acoust. *14:* 350–357 (1995).

O'Connor, J.D.; Arnold, G.F.: Intonation of colloquial English (Longmans, London 1961).

Pierrehumbert, J.: The phonology and phonetics of English intonation (MIT Press, Cambridge 1980).

Pierrehumbert, J.; Beckman, M.: Japanese tone structure (MIT Press, Cambridge 1988).

Pike, K.L.: Tone languages (University of Michigan Press, Ann Arbor 1948).

Prieto, P.; Santen, J. v.; Hirschberg, J.: Tonal alignment patterns in Spanish. J. Phonet. *23:* 429–451 (1995).
Santen, J.P.H. v.; Hirschberg, J.: Segmental effects on timing and height of pitch contours. Proc. Int. Conf. Spoken Lang. Processing, vol. 94: pp. 719–722 (Yokohama, 1994).
Shih, C.-L.: Tone and intonation in Mandarin. Working Papers, Cornell Phonetics Laboratory (No. 3): 83–109 (1988).
Silverman, K.E.A.; Pierrehumbert, J.B.: The timing of prenuclear high accents in English; in Kingston, Beckman, Papers in laboratory phonology 1 – between the grammar and physics of speech, pp. 72–106 (Cambridge University Press, Cambridge 1990).
Steele, S.A.: Nuclear accent F_0 peak location: effects of rate, vowel and number of following syllables. J acoust. Soc. Am. *80:* S51 (1986).
't Hart, J.; Cohen, A.: Intonation by rule: a perceptual quest. J. Phonet. *1:* 309–327 (1973).
Wang, W.S.Y.: Phonological features of tone. Int. J. Am. Ling. *33:* 93–105 (1967).
Woo, N.: Prosody and phonology; PhD diss. Massachusetts Institute of Technology (1969).
Xu, Y.: Putonghua yinlian de shengxue yuyinxue texing (Acoustic-phonetic characteristics of junctures in Mandarin). Zhongguo Yuwen: 353–360 (1986).
Xu, Y.: Yinjie he yinlian (Syllables and junctures); in Wu, Lin, An outline of experimental phonetics, pp. 193–220 (Higher Education Press, Beijing 1989).
Xu, Y.: Production and perception of coarticulated tones. J. acoust. Soc. Am. *95:* 2240–2253 (1994).
Xu, Y.: Contextual tonal variations in Mandarin. J. Phonet. *25:* 61–83 (1997).
Xu, Y.: Effects of tone and focus on the formation and alignment of F_0 contours – evidence from Mandarin (forthcoming: submitted to J. Phonet. and under revision).