Journal of the Internation Phonetic Association 33: 165-181 (author version)

Effects of Consonant Aspiration on Mandarin Tones

Ching X. Xu

Department of Communication Science and Disorders
Northwestern University
xxq@northwestern.edu

Yi Xu
Department of Linguistics,
The University of Chicago
xuyi@uchicago.edu

Abstract The influence of consonant aspiration on the F_0 contours of tones in Mandarin Chinese was tested in continuous speech by reference to a minimal pair of syllables, /ta/ and /tʰa/, which differ only in terms of aspiration. It was found that, consonant aspiration affects the fundamental frequency (F_0) of the following vowel. The onset F_0 of a tone is higher following unaspirated consonants than following aspirated consonants, and the magnitude of the differences are related to the tone itself as well as the preceding tone. The underlying mechanisms of this effect, as well as its interaction with other effects on F_0 contours, are discussed.

1. Introduction

Aspiration is an important distinctive feature of consonants in many languages. It divides stops and affricates into two groups: aspirated and unaspirated, which are associated with different phonemes in speech. When the aspiration interval is included, aspirated consonants are significantly longer than corresponding unaspirated ones (Feng 1985, Wu 1992), and the voice onset time (VOT) of an aspirated obstruents is longer than that of the unaspirated counterpart (Lisker & Abramson 1964, Ohde 1967 & 1984). Consonant aspiration affects not only the consonant itself, but also its adjacent vowels. A vowel is found to be shorter when it follows an aspirated consonant than following an unaspirated cognate, unless the aspiration is regarded as a part of the vowel (Feng 1985, Peterson & Lehiste 1960). This difference in vowel length seems to be due to a compensation effect for the consonant duration. However, the compensation is incomplete as the syllable containing an aspirated consonant is still longer than a syllable containing an unaspirated consonant (Feng 1985).

While the temporal effects of aspiration on consonants and vowels are widely accepted, there is no consensus regarding the effects of aspiration on the fundamental frequency (F_0) of adjacent vowels. Since aspiration seems to imply higher transglottal pressure with which F_0 varies proportionally, it is often assumed that an aspirated stop gives rise to a higher F_0 at the onset of the following vowel than an unaspirated cognate (Hombert 1975). This assumption is on the one hand supported by some data from various languages such as Danish (Jeel 1975), Korean (Han 1967, Han & Weitzman 1970, Kim 1968), Cantonese (Zee 1980), Thai and Japanese (Ewan 1976). However, it is on the other hand refuted by several studies of the same as well as other languages. For instance, Fischer-Jørgensen (1968) found no difference in F_0 values at the vowel onset regardless of the preceding consonants in Danish. Moreover, Gandour (1974) reported that in Thai, the onset F_0 of a vowel is slightly lower when the vowel follows an aspirated stop than an unaspirated counterpart. Similar results were also obtained in Hindi (Kagaya & Hirose 1975) and English (Ohde 1984, 1985), with even greater differences between the two conditions. Further counterevidence to the assumption is found in a study of Mandarin Chinese (Howie 1974). Howie illustrated tone variations between syllables with aspirated stops and unaspirated stops. Although he was mainly concerned with the overall pattern of F_0 contours and did not specify the F_0 onset values, his graphs depicted a lower onset F_0 following aspirated stops. As the measurement of F_0 during the first few cycles is usually hard, it is possible that the above conflicting data are due to different strategies of data extraction in different studies. Nevertheless, studies in Korean (Kagaya 1974) and Thai (Erickson 1975) both revealed that the results could be opposite across subjects within the same study.

In order to address this unresolved issue, Hombert & Ladefoged (1977) and Hombert (1978) investigated English voiceless aspirated stops as opposed to French voiceless unaspirated stops. They found that these two series of voiceless consonants had very similar effects on the F_0 contours of the following vowel. Nevertheless, this conclusion was based on indirect rather than direct comparisons. The involvement of subject and language factors rules out the possibility of making any direct comparisons between aspirated and unaspirated stops. Each set of voiceless stops was compared with their voiced cognates produced by the same speaker in the same language. The patterns of voiceless/voiced difference in the two languages were then matched qualitatively. Leaving alone the fact that voiced consonants in English and French may not be identical enough to serve as good references, the target voiceless stops are not well paired with regard to aspiration, as implied by the VOT values of French voiceless stops which are longer than one would expect for unaspirated stops (Hombert & Ladefoged 1977, Hombert 1978).

To provide new data and help clarify the effects of consonant aspiration on the following F_0, we conducted a study on Mandarin Chinese. Mandarin tones have relatively stable F_0 patterns, which offer good references for studying other effects on surface F_0 contours. More importantly, aspiration is a distinctive feature in Mandarin. There are 17 voiceless consonants in Mandarin, and 12 out of them, 6 stops and 6 affricates, are paired by the aspirated/unaspirated distinction. The six pairs are:

/p/ - /p^h/   /t/ - /t^h/   /k/ - /k^h/   /t□/ - /t□^{h}/   /ts/ - /ts^h/   /t□/ - /t□^{h}/

The two consonants in each pair are identical in all features except aspiration. Therefore, if there is any difference between the two sounds or the vowels following them, it is probably attributable or at least related to the effects of consonant aspiration.

Our concerns actually extend beyond how aspiration affects the following F_0. We are also interested in determining the source of the effect. Presumably the aerodynamic conditions related to aspirated and unaspirated consonants are different, and may result in different F_0 values at the vowel onset. The respiratory system usually generates a constant subglottal pressure (Ps) during closure for all stops (Ladefoged 1967, Löfqvist 1975, Ohala & Ohala 1972, Slis 1970). At the release of an aspirated stop, a high rate of airflow runs through the glottis and Ps decreases markedly during the aspiration period. At the release of an unaspirated stop, in contrast, little air flows out of the subglottal area and Ps diminishes fairly gradually. By the time of vowel onset, therefore, Ps is generally lower after an aspirated stop than an unaspirated one (Ladefoged 1963 & 1974, Ohala & Ohala 1972, Ohala 1978), as described in a speech aerodynamic model proposed by Ohala (1975, 1976 & 1978). Since the oral pressure after a stop release is negligible, lower Ps means a lower transglottal pressure, which in turn should lead to a relatively lower F_0. Aspirated stops thus should give rise to lower onset F_0 of the following vowel than unaspirated cognates. This prediction is at odds with the assumption of Hombert (1975) and some research data (e.g. Jeel 1975), but is in agreement with most other experimental results (e.g. Gandour 1974).

A second source of possible F_0 variations relevant to consonant may relate to the state of vocal folds. It is known that during the interval of voiceless obstruents, the vocal folds are stiffened in order to prevent vocal-fold vibration (Stevens 1998, Hanson & Stevens, 2002). The stiffening action can carry over into the following vowel and lead to a raising of F_0 at the vowel onset. Nevertheless, this should be true for both voiceless aspirated consonants and voiceless unaspirated consonants. Moreover, there has been no evidence of different muscular activity around the glottis after aspirated as opposed to unaspirated stops when the following vowels start (Hombert & Ladefoged 1977, Hombert 1978). Figure 1 illustrates a hypothetical schematic timing of related parameters during the production of aspirated and unaspirated stops. From top to bottom, the hypothetical parameters are: articulatory gestures of oral cavity and glottis, corresponding acoustic excitations and Ps variations, respectively. Both the voiceless aspirated stop (on the left) and the voiceless unaspirated stop (on the right) consist of two shaded intervals in the graph. The first shades correspond to the stop closure, during which Ps builds up to a constant level. The second shades correspond to voice onset time (VOT), which lasts from stop release to vowel onset. As shown in the graph, the state of glottis varies differently during the consonant interval across stops, but by the time voicing starts, the glottis has been closed no matter which stop precedes the vowel. In contrast, although it builds up to the same level during closure for both aspirated and unaspirated stops, Ps would reach a lower value following an aspirated stop than following an unaspirated stop at the vowel onset. Taken together, these considerations suggest that aspirated consonants are associated with lower F_0 at the onset of the following vowel than unaspirated cognates, and this difference should be due to an aerodynamic effect.

Insert Figure 1 about here

Since the consonant aspiration effect on F_0 seems nontrivial, a further issue is how the effect interacts with F_0 variations originating from other factors. In particular, tone information, which is phonemic in Mandarin, is mainly conveyed by the variations of F_0 contours. The relationship between implementation of lexical tone and F_0 perturbation by consonants has yet to be explored. According to a pitch target approximation model we have proposed (Xu, Xu & Luo 1999, Xu & Wang 2001), the basic units underlying Mandarin tones are pitch targets that are either static or dynamic, and each pitch target is implemented continuously and asymptotically in synchrony with a syllable (Xu 1998), as illustrated in Figure 2. Considering the fact that tone carries lexical information and is unchanged across syllables that are different with respect to segments, its underlying pitch target should remain intact despite F_0 perturbations by consonants. Nevertheless, the implementation of the pitch target is likely to be influenced by the consonant effect. First of all, the starting F_0 varies across syllables with different consonants. Approximation of the same target with different initial values necessitates different approaching routes. Secondly, the timing of F_0 contours varies across syllables with different consonants. With the presence of voiceless intervals, voicing is interrupted during the consonant. Thus the visible F_0 trajectory of a syllable starts from vowel onset instead of syllable onset. Because a vowel is shorter after an aspirated consonant than after an unaspirated cognate (Feng 1985), syllables with aspirated consonants carry shorter F_0 curves than syllables with unaspirated consonants. This temporal difference may make the tonal contours appear to approach the underlying pitch targets sooner following aspirated consonants than following unaspirated ones. Consequently, although the aerodynamic effect lasts for only a few milliseconds after voicing begins (Ohala, 1978), the later parts of F_0 contours following two different consonants will not fully converge if they are aligned with the vowel onset. Rather, if the curves are aligned with the vowel offset, which is also the offset of the syllable as well as the tone it carries, they should show more similarities with each other during the process of approximating the underlying pitch target.

Insert Figure 2 about here

While the effect of consonant aspiration on F_0 may influence tone implementation, as discussed above, it itself may be influenced by tone. This is due to the quantitative relationship between F_0 and Ps (van den Berg 1957, Ladefoged 1963, Öhman and Lindqvist 1968, Lieberman et al. 1969, Hixon et al. 1971, Baer 1979, Rothenberg & Mahshie 1986), which varies with the state of vocal folds. When the intended pitch is high within the chest register, the vocal folds become tense and their oscillation is less likely to be affected by aerodynamic factors. When the intended pitch is low, in contrast, vocal folds become lax and their activities are more likely to be affected by aerodynamic factors. Therefore, the change of F_0 with Ps is smaller for high pitch than for low pitch, as suggested by Atkinson (1978) and Titze (1989a). Since the intended onset pitch is high for the High and Falling Tones and low for the Rising and Low Tones, F_0 differences at the vowel onset following aspirated versus unaspirated consonants should be greater for the Rising and Low Tones than for the High and Falling Tones.

Contextual tonal variation is another factor that often contributes to shape surface F_0 curves in continuous speech (Chao 1968, Xu 1997 & 2001). Figure 3 sketches two examples of F_0 implementation of two successive tones in the framework of the pitch target approximation model. Here voicing continues across syllable boundaries, so that the onset F_0 of a tone is determined by the offset F_0 of its preceding tone (Xu 1997). Accordingly, the F_0 contour of the second tone varies with the offset F_0 of the first tone. This carryover effect is assimilatory, and its magnitude decreases over time, as shown in Figure 4, which is adapted from Xu (1997). By comparing the curves in the four panels of Figure 4, we can also see a subtle anticipatory effect that is dissimilatory: a Low Tone tends to raise the F_0 contour of its preceding tone (Gandour et al 1992, Laniran & Clements 1995, Xu 1993, 1997). When a voiceless consonant is present and interrupts continuous voicing, these contextual tonal variations may interact with the F_0 perturbations due to aspirated as well as those due to unaspirated consonants. Comprehension of this interaction may not only lead to a better understanding of the resulting F_0 trajectory, but also shed further light on the underlying mechanisms of each individual effect.

Insert Figure 3 about here

Insert Figure 4 about here

In order to address the above issues, the present study investigated the F_0 variations relating to voiceless initial aspirated versus unaspirated stops in continuous Mandarin speech. Based on the results, we will have further discussion on our hypotheses and concerns.

2. Experiment

2.1. Stimuli

The stimuli were syllables /ma/, /ta/, /t^{h}a/ and /□a/. Written in the Mandarin Pinyin system, they are *ma*, *da*, *ta* and *sha*, respectively. /ta/ and /t^{h}a/ were the real targets of investigation in this study. /ma/ and /□a/ were for other uses, while also serving as fillers for the present experiment. These four syllables can carry any of the four Mandarin tones, (here symbolized by the numerals 1, 2, 3 and 4 representing High, Rising, Low and Falling Tone, respectively), except that /t^{h}a/ with Rising Tone, i.e. /t^{h}a2/, is lexically missing in Mandarin. According to Ohde (1984), /t^{h}a/ and /p^{h}a/ are consistent in terms of F_0 patterns, so /p^{h}a2/ was used as a substitute.

The target tonal syllables were combined into phonetically legal disyllabic words. These words were sorted into two lists. In one list the target syllables were in the first position and in the other the target syllables were in the second position. In both lists, the target tonal syllables had all possible tonal contexts (i.e. four tones). Only the sequence Low-Low was excluded since it is not phonetically different from Rising-Low due to a phonological transformation known as tone sandhi (Chao, 1968).

The disyllabic words were incorporated into two carrier sentences, *wo3 lai2 shuo1* ____ *zhe4 ge4 ci2* ("I say the word ____") and *wo3 lai2 zhao3* ____ *zhe4 ge4 ci2* ("I look for the word ____"). Only the syllables preceding the target words were different in the two conditions. The first had a High Tone, while the second had a Low Tone.

2.2. Subjects

Seven female native speakers of Mandarin participated as subjects. They were recruited from the Northwestern University community. All subjects had lived in the United States for approximately two years at the time of recording. They had no history of speech and language impairments. Their ages ranged from 22 to 30 years old.

2.3. Procedure

The recording was made in a soundproof booth, one subject per session. The subject was seated comfortably in front of a computer screen and wore a head-mount microphone approximately 2 inches away from her mouth.

At the beginning of the session, a set of instructions was displayed on the screen, directing the subject to read the target sentences aloud at a normal speaking rate and with stress on the target words. After reading the instructions, the subject went through a series of practice trials to become familiar with the target words before the real trials. In each trial, an experimental sentence in Chinese characters appeared on the screen, and the subject read aloud the sentence as instructed. The subject had control of two buttons on the screen: "Next" and "Again". Thus, she could choose either to continue with the next sentence when she had successfully finished the current one, or to repeat the current sentence if she had made a mistake.

Each experimental sentence was presented to the subject five times and the order of the repetitions was randomized. There were a total of 1110 sentences (111 target words x 2 carrier sentences x 5 repetitions). Each session lasted for about one hour. The subject's speech was digitized by SoundEdit (Macromedia Inc.), with 22 kHz sampling rate and 16-bit accuracy.

2.4. Data Extraction and Measurement

F_0 curves of target syllables were extracted using a method that combined automatic vocal cycle detection and manual rectification (Xu 1997). Using the ESPS signal processing software package (Entropic Inc.), every vocal cycle in the target words was marked by the program EPOCHS. The marks were manually checked and corrected. The first period of each target syllable was thrown out if it was apparently not a full cycle. At the same time, the onset and offset of each segment of the target syllables were manually labeled using the program XLABEL in XWAVES.

The markings of vocal pulses between the segment labels were processed to extract the F_0 value of each vocal cycle. The F_0 trajectories were smoothed individually by applying a three-point median filter. Finally, the five repetitions of each syllable by each subject were averaged to remove random variations.

3. Analyses and Results

Since there are five factors (consonant, lexical tone, syllable position, tonal context and carrier sentence) involved in the experiment, we performed five-way ANOVA tests in SPSS for statistical analyses. The outcomes for the seven subjects show the same pattern of F_0 variations. For the sake of data presentation, we report results only of pooled analyses, which use means across the five repetitions of all subjects as the dependant variables.

Table 1 lists the most relevant results of an ANOVA test on the onset F_0 values (i.e. the F_0 values of the first valid vocal cycles) of target syllables. The main effect of consonant aspiration and lexical tone, as well as their interactions are all significant. Post hoc LSD test of tone reveals that each of the four tones is significantly different from one another. The significant interaction between syllable position and tonal context reflects the different nature of carryover and anticipatory effects. The three-way interaction of consonant aspiration, syllable position and tonal context indicates that either carryover or anticipatory effect interact with the consonant aspiration effect, as will be further explained and illustrated in the following sections. The effect of the carrier sentence is also significant. It is a carryover effect across the word boundary, which is less significant but fairly similar to the carryover effect within a word. As a result, the effect of the carrier sentence is neither listed in the table nor discussed in more detail in the following. Rather, it should be considered as included when we refer to carryover effect.

Insert Table 1 about here

To statistically analyze the F_0 patterns towards syllable offset, data were extracted from F_0 curves every 5 milliseconds from the offset by interpolations, which were then subjected to ANOVA tests. The main effect of consonant aspiration is not significant at $\alpha=0.05$ level for the first 21 points, i.e. within 105 milliseconds from the syllable offset. The main effect of lexical tone remains significant all the way through, and post hoc tests indicate that there are overlaps and crossings of the four tones, as will be shown in the following graphic illustrations.

3.1. Main Effect of Aspiration on F_0 and Its Interaction with Lexical Tone

The main effect of consonant aspiration on onset F_0 value of a syllable is shown in Figure 5. Both columns represent averaged values over all other factors and the horizontal bars indicate standard error. Evidently, the F_0 value at the voice onset is significantly higher for /ta/ than for /t^{h}a/.

Insert Figure 5 about here

The interaction of consonant aspiration with lexical tone on onset F_0 of a syllable is shown in Figure 6. Each column represents an average over the factors other than consonant and lexical tone. Horizontal bars indicate standard error. Consistent with Figure 5, the F_0 values at the voice onset are higher for /ta/ than for /t^{h}a/ across the four tone conditions. However, the magnitude of the difference varies with tone. When the tone is Rising or Low, the difference between /ta/ and /t^{h}a/ is much greater than when the tone is High or Falling. The main effect of lexical tone can also be seen in Figure 6. For both /ta/ and /t^{h}a/, the F_0 onset is much higher for High and Falling Tone than for Rising and Low Tone.

Insert Figure 6 about here

Effects of consonant aspiration on the F_0 contours of four tones are illustrated in Figure 7. Each curve represents an average across 5 repetitions, 4 posterior tonal contexts and 7 subjects. The syllables averaged were produced in the first position of disyllabic stimulus words and with a high preceding tone in the carrier sentence. All curves are aligned to the syllable offset.

Insert Figure 7 about here

The F_0 differences between /ta/ and /t^{h}a/ at the voice onset are consistent with what were shown in Figure 6. Interestingly, no matter how big the differences are, towards the syllable offset the two F_0 curves with the same tone converge to the same pattern. For High Tone and Low Tone, the patterns are level high and low, respectively. For Rising and Falling Tone, the patterns are sloping rising and falling, respectively.

3.2. Interaction of Aspiration Effect with Tonal Coarticulation

The interaction of consonant aspiration with preceding tonal context on the onset F_0 of a syllable is shown in Figure 8. Each column represents an average over 5 repetitions, 2 carrier sentences, 4 lexical tones and 7 subjects. All target syllables were produced in the second position of the disyllabic words. Horizontal bars indicate standard error.

Insert Figure 8 about here

The magnitude of the effect of consonant aspiration on F_0 seems to vary with the preceding tones. When the preceding tone is High or Rising, the F_0 onset is much higher for /ta/ than for /t^{h}a/. When the previous tone is Low or Falling, in contrast, the F_0 onsets of the two syllables do not differ as significantly. Apparently, the kind of assimilatory carryover effect as reported in previous studies (Xu 1997, 1999) holds with the presence of voiceless consonants. After High or Rising Tone, the F_0 contours of both /ta/ and /t^{h}a/ start at higher values than after Low or Falling Tone. Nevertheless, the magnitude of carryover effect at the voice onset is much greater for /ta/ than for /t^{h}a/.

Figures 9 and 10 illustrate the relationship between the effects of consonant aspiration and carryover tonal influence on the F_0 contours in two forms. Each curve represents an average across 5 repetitions, 2 carrier sentences and 7 subjects. All target syllables carry the Rising Tone and were produced in the second position of the disyllabic words. As in Figure 7, all curves are aligned to the syllable offset.

Insert Figure 9 about here

Insert Figure 10 about here

Plotting together the F_0 contours of /ta/ and /t^ha/ under the same condition in each panel, Figure 9 shows the same results as in Figure 8 regarding the onset F_0 values. Besides, it further demonstrates that the F_0 curves with the same tone converge to the same pattern towards the syllable offset. In the current example, the pattern is a sloping rise as the target syllables carry the Rising Tone. Figure 10 plots the F_0 contours with different preceding tones together, and demonstrates that the F_0 variations due to previous tones decrease over time and almost vanish before the syllable offset.

The interaction of consonant aspiration with the following tonal context on the onset F_0 of a syllable is shown in Figure 11. Each column represents an average over 5 repetitions, 2 carrier sentences, 4 lexical tones and 7 subjects. All target syllables are produced in the first position of the disyllabic words. Horizontal bars indicate standard error.

Insert Figure 11 about here

The differences of onset F_0 values between /ta/ and /t^ha/ are fairly consistent across the four conditions in Figure 11. Interestingly, the onset F0 values are slightly higher before the Low Tone than before other tones. The phenomenon seems like a reflection of dissimilatory anticipatory effect (Gandour et al., 1992; Xu, 1997).

To see the relationship of consonant aspiration effect and tonal anticipatory effect more clearly, Figures 12 and 13 illustrate the interaction between the two on the F_0 contours of syllables. Each curve represents an average across 5 repetitions, 2 carrier sentences and 7 subjects. All target syllables carry the Rising Tone and are produced in the first position of the disyllabic words. Again, all curves are aligned to the syllable offset. Plotting together the F_0 contours of /ta/ and /t^ha/ under the same condition in each panel, Figure 12 demonstrates the same aspiration effect on the onset F_0 as shown in Figure 11 and the same pattern of Rising Tone as shown in Figure 9. Nevertheless, the exact slope of the rising pattern, as well as the

Insert Figure 12 about here

Insert Figure 13 about here

onset F_0 values of /ta/ and /t^ha/, may vary with the following tone because of the anticipatory effect, as shown in Figure 13 where the F_0 contours with different following tones are plotted together.

4. Discussion

The results confirm our hypotheses that the effect of consonant aspiration on the following F_0 is significant, and the effect is greater for the Rising and Low Tones than for the

High and Falling Tones. Nevertheless, the aspiration effect may originate not only from changes of subglottal pressure, but also from variations of vocal folds state, as the main effect of lexical tone on onset F_0 value would suggest. Previous studies have shown that onset F_0 values of voiced syllables are determined by offset F_0 values of the preceding tones (Xu 1997, 1999, Xu, Xu & Luo 1999). The present target syllables, which contain voiceless initial consonants, may appear at first glance not to directly inherit F_0 from the preceding tones. Although they display the same trend of assimilated carryover effects, their onset F_0 values are more closely related to the current tones, as demonstrated by the statistical analyses. The close relationship between onset F_0 and lexical tone implies that the state of vocal folds is being adjusted to implement the current tone during the voiceless consonants. By the time voicing starts, there have been substantial laryngeal adjustments towards the goal. As a result, when the visible F_0 contours appear, although the influence of the preceding tone is still present, the F_0 value already more closely resemble that of the current tone. This inference is supported by the fact that the magnitude of carryover effect at the voice onset is greater for /ta/ than for /t^{h}a/. Since unaspirated consonants are shorter than the aspirated cognates (Feng 1985, Wu 1992), /ta/ has provided less time for the adjustments towards the tonal target than /t^{h}a/ by the time voicing starts. As a result, the onset F_0 is more influenced by previous tonal offsets in /ta/ than in /t^{h}a/, but is more closely related to current tonal targets in /ta/ than in /t^{h}a/. Naturally, physiological studies are needed to provide further evidence in terms of muscle activities and vocal fold tension around aspirated and unaspirated stops as the current and surrounding tones vary.

The contribution of aerodynamics on F_0 values can be more clearly seen in cases where the vocal folds should be in similar states right after aspirated versus unaspirated consonants. This is true when a High Tone is preceded by another High Tone, as the state of vocal folds is supposed to change little (other than the momentary tension increase due to devoicing (Stevens 1998, Hanson & Stevens, 2002)) during the transition from the first tone to the second tone, as well as throughout the second tone. Then the F_0 difference due to aspiration in this case should be mainly, if not totally, attributable to the aerodynamic effect. The corresponding F_0 curves of /ta/ versus /t^ha/ are illustrated in Figure 14. Each curve represents an average across 5 repetitions, 2 carrier sentences and 7 subjects. All target syllables were produced in the second position of the disyllabic words. Statistical analysis shows that the onset F_0 values of the two curves are significantly different ($P < .001$). This result implies the significance of aerodynamic effect, and in turn lends support to our hypothesis regarding variations of F_0 with Ps. During the closure of the stops, Ps builds up to a constant level irrespective of the aspiration feature of the consonants. At the release of /t^h/, Ps decreases markedly because of the rapid outflow of air, and then gradually builds back up at vowel onset. At the release of /t/, however, Ps remains at a high level and then gradually returns to normal. Therefore, Ps should be much lower at the voice onset for /t^ha/ than for /ta/. The oral pressure should be very low in both series after the release. If there is any difference, it would be that the oral pressure is higher after /t^h/, since the higher flow rate will cause a higher back pressure (Hombert & Ladefoged 1977). Other things being equal, these differences should lead to lower onset F_0 in /t^ha/ than in /ta/.

Insert Figure 14 about here

It may be noticed from Figure 12 that the F_0 variations related to Ps is around 20Hz, which is higher than one would expect normally (Titze 1989a). This can probably be explained by considering the state of the vocal folds. It is well known that, male voices and female voices differ both physiologically and acoustically (Titze 1989b). Females generally have shorter and less massive vocal folds than males, which make them more sensitive to aerodynamic changes in speech. Since most of the well-established facts regarding F_0 variations with Ps come from studies on male speakers, it is not surprising that the aerodynamic effects with female subjects are higher than what have been reported.

When aligned with the syllable offset, the F_0 curves of /ta/ and /t^{h}a/ with the same tone appear to converge long before reaching the syllable offset, after which they follow the same course all the way through. The robust F_0 pattern in each tone condition gives further support to our assumption that there is an underlying pitch target for each tone (Xu, Xu & Luo 1999, Xu, 1998, Xu & Wang 2001). Presumably, the targets are static [high] and [low] for the High Tone and Low Tone, respectively, and dynamic [rise] and [fall] for the Rising Tone and Falling Tone, respectively. The aspiration effect on F_0 does not change the underlying pitch targets. It only superimposes local F_0 variations at the vowel onset. The evidence that there are laryngeal adjustments towards tone implementation during the interval of voiceless consonants is consistent with what Xu (1998) found about the tone-syllable alignment. It lends further support to our assumption in the pitch target approximation model that a pitch target is implemented in synchrony with the entire syllable rather than with just the voiced portion of the syllable.

Consequently, contextual tonal variations are maintained with the presence of voiceless stops. The carryover effect is greatest at the voice onset and then decreases over time towards the syllable offset. As the pause of voicing gives time for some laryngeal adjustments, the carryover effect at the voice onset for both /ta/ and /t^{h}a/ may not seem as great as those reported in Xu (1997), which uses a nasal as the initial consonant. As our present data have demonstrated, the effect is only partially hidden rather than totally missing. The anticipatory effect is consistent with the previous finding (Xu 1997), and it seems to influence slightly the onset F_0 as well as the whole F_0 contours. In general, therefore, the aspiration effect is superimposed onto the F_0 contours resulting from the contextual tonal variations.

5. Conclusions

Consonant aspiration in Mandarin has a significant effect on the following onset F_0, with aspirated stops associated with lower values and their unaspirated cognates with higher values. This effect is attributable to aerodynamic factors, as well as myoelastic factors. Being local, the aspiration effect does not change any of the underlying pitch targets of lexical tones. Rather, it is superimposed onto the F_0 contours resulting from known contextual tonal variations without effectively altering the general course of the F_0 movements.

6. Acknowledgements

The authors would like to thank Ciara Leydon, Peter Ladefoged, Charles Larson and Bruce Smith, as well as two anonymous reviewers for their valuable comments on an earlier version of the paper. Part of the study was presented at 142nd Meeting of the Acoustical Society of America (Xu & Xu 2001).

7. References

• ATKINSON, J.E. (1978) Correlation analysis of the physiological factors controlling fundamen-tal voice frequency. *J. Acoust. Soc. Am.* 63, 211-222.

• BAER, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. *J. Acoust. Soc. Am.* 65, 1271-1275.

• CHAO, Y.R. (1968). *A grammar of Spoken Chinese*. Berkeley, CA: Univ. of California Press.

- ERICKSON, D. (1975). Phonetic implication for a historical account of tonogenesis in Thai. In Harris, J.G. & Chamberlain, J.R. (Eds.), *Studies in Tai Linguistics in Honor of W. J. Gedney*, 110-111. Bangkok: Central Institute of English Language Office of State Universities.
- EWAN, W.G. (1976). *Laryngeal Behavior in Speech*. Ph.D. dissertation, Univ. of California, Berkeley.
- FENG, L. (1985). Duration of initials, finals and tones in Beijing dialect. *Working Papers in Experimental Phonetics*, 131-195. Peking Univ. Press, Beijing.
- FISCHER JØRGENSEN, E. (1968). Les occlusives françaises et danoises d'un sujet bilingue. *Word*, 24, 112-153.
- GANDOUR, J.T. (1974). Consonant types and tone in Siamese. *J. Phonet*. 2, 337-350.
- GANDOUR, J.T., POTISUK, S., DECHONGKIT, S. & PONGLORPISIT, S. (1992). Anticipatory tonal coarticulation in Thai noun compounds. *Linguistics of the Tibeto-Burman Area*, 15, 111-124.
- HAN, M.S. (1967). Acoustic features in the manner differentiation of Korean stop consonants. *Studies in the Phonology of Asian Languages*, vol. 5. Los Angeles: Acoustic Phonetics Research Lab, Univ. of Southern California.
- HAN, M.S. & WEITZMAN, R.S. (1970) Acoustic features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/. *Phonetica* 22, 112-128.
- HANSON, H.M. & STEVENS, K.N. (2002) A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *J. Acoust. Soc. Am*. 112 (3), Pt. 1, 1158-1182.

- HIXON, T.J., KLATT, D. H. and MEAD, J. (1971) Influence of forced transglottal pressure on fundamental frequency. *J. Acoust. Soc. Am.* 49, 105(A).
- HOMBERT, J.M. (1975). *Towards a theory of tonogenesis: An empirical, physiologically, and perceptually-based account of the development of tonal contrasts in language.* Unpublished Ph.D. dissertation, Univ. of California, Berkeley.
- HOMBERT, J.M. (1978). Consonant types, vowel quality, and tone. In V.A. Fromkin (Ed.), *Tone: A Linguistic Survey*, 77-111. Academic Press, New York.
- HOMBERT, J.M. & LADEFOGED, P. (1977). The effect of aspiration on the fundamental frequency of the following vowel. *UCLA Work. Pap. Phonet.* 36, 33-40.
- HOWIE, J.M. (1974) On the domain of tone in Mandarin. *Phonetica*, 30, 129-148.
- ISSHIKI, N. & RINGEL, R. (1964). Air flow during the production of selected consonants. *J. Speech Hear. Res.* 7, 233-244.
- JEEL, V. (1975). An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. *Annual Report of the Institute of Phonetics, Univ. of Copenhagen*, 9, 191-211.
- KAGAYA, R. (1974). A fiberoptic and acoustic study of the Korean stops, affricates, and fricatives. *J. Phon.* 2(2), 161-180.
- KAGAYA, R. & HIROSE, H. (1975) Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 9, 27-46.
- KIM, C. W. (1968). Review of Lieberman 1967. *Lang.* 44, 830-842.

- LADEFOGED, P. (1963). Some physiological parameters in speech. *Lang. Speech* 6, 109-119.
- LADEFOGED, P. (1967). *Three areas of experimental phonetics*. London: Oxford Univ. Press.
- LANIRAN, Y.O. & CLEMENTS, G.N. (1995). A long-distance dependency in Yoruba tone realization. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 2, 734-737, Stockholm.
- LIEBERMAN, P., KNUDSON, R. & MEAD, J. (1969). Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. *J. Acoust. Soc. Am.* 45, 1537-1543.
- LISKER, L. & ABRAMSON, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical Measurements. *Word*, 20, 384-422.
- LÖFQVIST, A. (1975). Intrinsic and extrinsic F_0 variations in Swedish tonal accents. *Phonetica*, 31, 228-247.
- OHALA, J.J. (1975). A mathematical model of speech aerodynamics. In G. Fant (Ed.), *Speech communication: Proceedings of the Speech Communication Seminar, Stockholm, August 1-3 1974 (vol. 2): Speech Production and Synthesis by Rule*, 65-72. Stockholm: Almqvist & Wiksell.
- OHALA, J.J. (1976). A model of speech aerodynamics. *Report of the Phonology Laboratory*, Berkeley, 1, 93-107.
- OHALA, J.J. (1978). The production of tone. In V.A. Fromkin (Ed.), *Tone: A Linguistic Survey*, 5-39. Academic Press, New York.

- OHALA, M. & OHALA, J.J. (1972). The problem of aspiration in Hindi phonetics. *Project on Linguistic Analysis Reports*, Berkeley, 16, 63-70; *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, 6, 39-46.
- OHDE, R. (1967). Some effects of context on voice onset time in English stops. *Lang. Speech* 10, 1-28.
- OHDE, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *J. Acoust. Soc. Am.* 75, 224-230.
- OHDE, R. (1985). Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children. *J. Acoust. Soc. Am.* 78, 1554-1561.
- ÖHMAN, S. & LINDQVIST, J. (1968). Analysis and synthesis of prosodic pitch contours. *Z. Phonet.* 21, 164-170.
- PETERSON, G.E. & LEHISTE, I. (1960). Duration of syllable nuclei in English. *J. Acoust. Soc. Am.* 32, 693-703.
- ROTHENBERG, M. & MAHSHIE, J. (1986). Induced transglottal pressure variations during voicing. *J. Phon.* 14, 365-371.
- SLIS, I.H. (1970). Articulatory measurements on voiced, voiceless, and nasal consonants. *Phonetica*, 21, 193-210.
- STEVENS, K.N. (1998) *Acoustic Phonetics*. MIT Press, Cambridge, MA.
- TITZE, I.R. (1989a). On the relation between subglottal pressure and fundamental frequency in phonation. *J. Acoust. Soc. Am.* 85(2), 901-906.
- TITZE, I.R. (1989b). Physiologic and acoustic differences between male and female voices. *J. Acoust. Soc. Am.* 85(4), 1699-1707.

- VAN DEN BERG, J. (1957). Sub-glottal pressure and vibrations of the vocal folds. *Folia Phoniatr*. 9, 65-71.
- VAN HATTUM, R.J. & WORTH, J.H. (1967). Air flow rates in normal speakers. *Cleft Palate Journal*, 4, 137-147.
- WU, Z. (1992). *Essentials of modern Chinese phonetics (in Chinese)*. Foreign Language Printing House, Beijing.
- XU, C.X., XU, Y. & LUO, L. (1999) A pitch target approximation model for F0 contours in Mandarin. *In Proc. of 14th Internat. Cong. – Phon. Sc.*, San Francisco, pp. 2359-2362.
- XU, C.X. & XU, Y. (2001). Effects of consonant aspiration on Mandarin tones. *J. Acoust. Soc. Am.* 110, 2655.
- XU, Y. (1993) *Contextual Tonal Variation in Mandarin Chinese*. Ph.D. dissertation, The University of Connecticut.
- XU, Y. (1997). Contextual tonal variations in Mandarin. *J. Phon.* 25, 61-83.
- XU, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179-203.
- XU, Y. (1999) Effects of tone and focus on the formation and alignment of F0 contours. *J. of Phon.* 27**,** 55-105.
- XU, Y. (2001) Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, Monograph series #17, 1-31.
- XU, Y. & WANG, E.Q. (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication*, 33: 319-337.

- ZEE, E. (1980). The effect of aspiration on the F0 of the following vowel in Cantonese. *UCLA Work. Pap. Phonet.* 49, 90-97.

A list of figure and table captions

Figure 1. Schematic timing of related parameters during Mandarin speech. From top to bottom, the parameters are: articulatory gestures of oral cavity and glottis, corresponding acoustic excitations and Ps variations, respectively. Both the voiceless aspirated stop (on the left) and the voiceless unaspirated stop (on the right) consist of two parts, that are shaded. The first marks the stop closure. The second indicates the voice onset time (VOT). (Part of the graph was courtesy of Anders Löfqvist. We added the hypothetical Ps tracing and the shading.)

Figure 2. A schematic illustration of the pitch target implementation model as applied to the four lexical tones in Mandarin. The vertical boundaries of each individual graph represent syllable onset and offset. The dashed lines represent underlying pitch targets. The solid lines represent the F_0 contours resulting from articulatory implementation of the pitch targets. The High and Low Tones have static [high] and [low] as targets, while the Rising and Falling Tones have dynamic [rise] and [fall] as targets, respectively.

Figure 3. A schematic illustration of the F_0 contours of two successive tones based on the pitch target implementation model. The vertical lines represent syllable boundaries. The dashed lines represent underlying pitch targets. The solid lines represent the F_0 contours resulting from articulatory implementation of the underlying pitch targets.

Figure 4. Effects of preceding tone on the F_0 contour of the following tone in Mandarin. In each panel, the tone of the second syllable is held constant, while the tone of the first syllable is either High, Rising, Low or Falling. The vertical lines represent the syllable boundaries (at the onsets of initial nasals). Each curve is a (segment-by-segment) time-normalized average of 192 tokens produced by eight speakers. (Adapted from Xu 1997)

Figure 5. Average onset F_0 values of /ta/ versus /t^{h}a/.

Figure 6. Average onset F_0 values of /ta/ versus /t^{h}a/ across four tones.

Figure 7. Average F_0 contours of syllable /ta/ versus /t^{h}a/ in four tones. The right boundary of each panel represents syllable offset.

Figure 8. Average onset F_0 values of /ta/ versus /t^{h}a/ with different preceding tones.

Figure 9. Average F_0 contours of /ta/ versus /t^{h}a/ with Rising Tone following different tonal contexts. The right boundary of each panel represents syllable offset.

Figure 10. Average F_0 contours of /ta/ and /t^{h}a/ with Rising Tone following different tones. The right boundary of each panel represents syllable offset.

Figure 11. Average onset F_0 values of /ta/ versus /t^{h}a/ with different following tones.

Figure 12. Average F_0 contours of /ta/ versus /t^{h}a/ with Rising Tone preceding different tonal contexts. The right boundary of each panel represents syllable offset.

Figure 13. Average F_0 contours of /ta/ and /t^{h}a/ with Rising Tone when followed by different tones. The right boundary of each panel represents syllable offset.

Figure 14. Average F_0 contours of /ta/ versus /t^{h}a/ of High Tone with High Tone preceding it.

Table 1. Results of an ANOVA test on the onset F_0 values of target syllables.

Table 1. Results of an ANOVA test on the onset F_0 values of target syllables.

Source of Effect	F	P
Consonant	60.20	< .001
Tone	179.6	< .001
Consonant × Tone	7.55	< .001
Position × Context	12.79	< .001
Consonant × Position × Context	3.15	.025

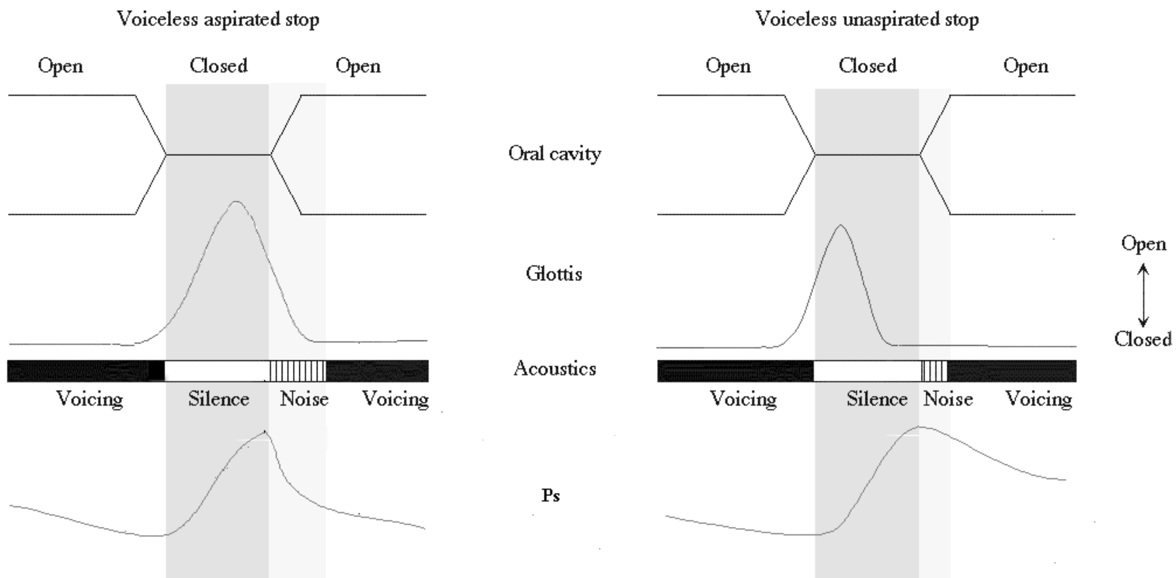

Figure 1. Schematic timing of related parameters during Mandarin speech. From top to bottom, the parameters are: articulatory gestures of oral cavity and glottis, corresponding acoustic excitations and Ps variations, respectively. Both the voiceless aspirated stop (on the left) and the voiceless unaspirated stop (on the right) consist of two parts, that are shaded. The first marks the stop closure. The second indicates the voice onset time (VOT). (Part of the graph was courtesy of Anders Löfqvist. We added the hypothetical Ps tracing and the shading.)

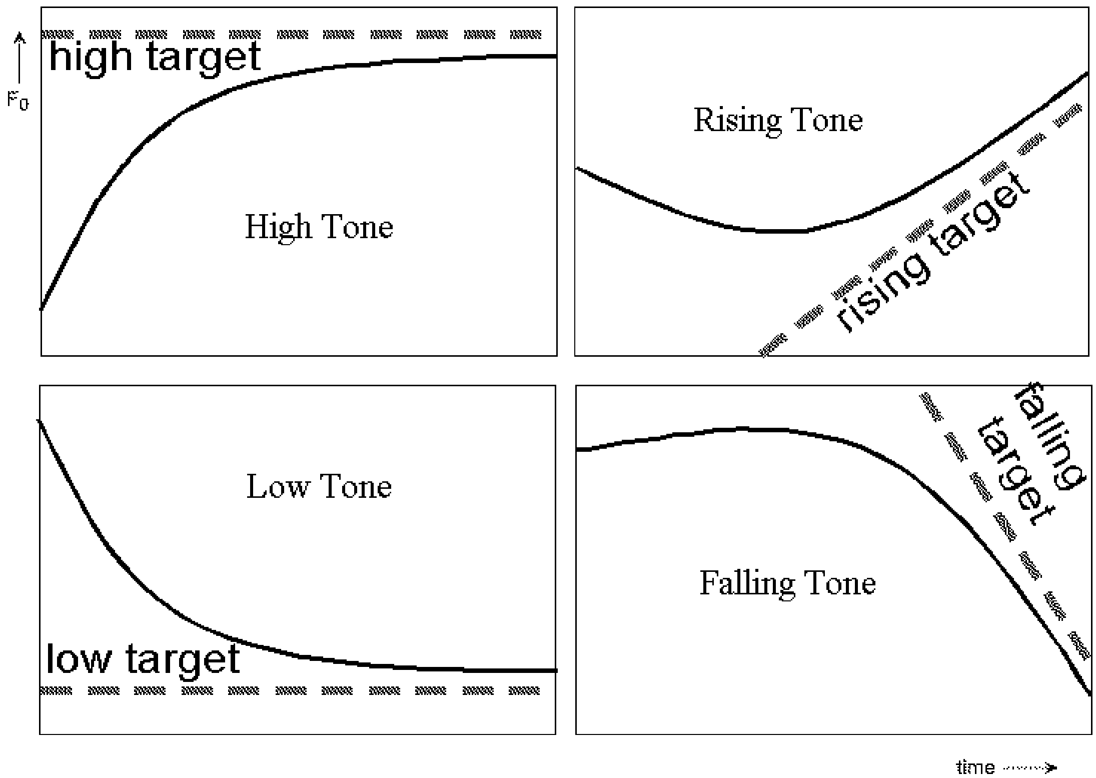

Figure 2. A schematic illustration of the pitch target implementation model as applied to the four lexical tones in Mandarin. The vertical boundaries of each individual graph represent syllable onset and offset. The dashed lines represent underlying pitch targets. The solid lines represent the F_0 contours resulting from articulatory implementation of the pitch targets. The High and Low Tones have static [high] and [low] as targets, while the Rising and Falling Tones have dynamic [rise] and [fall] as targets, respectively.

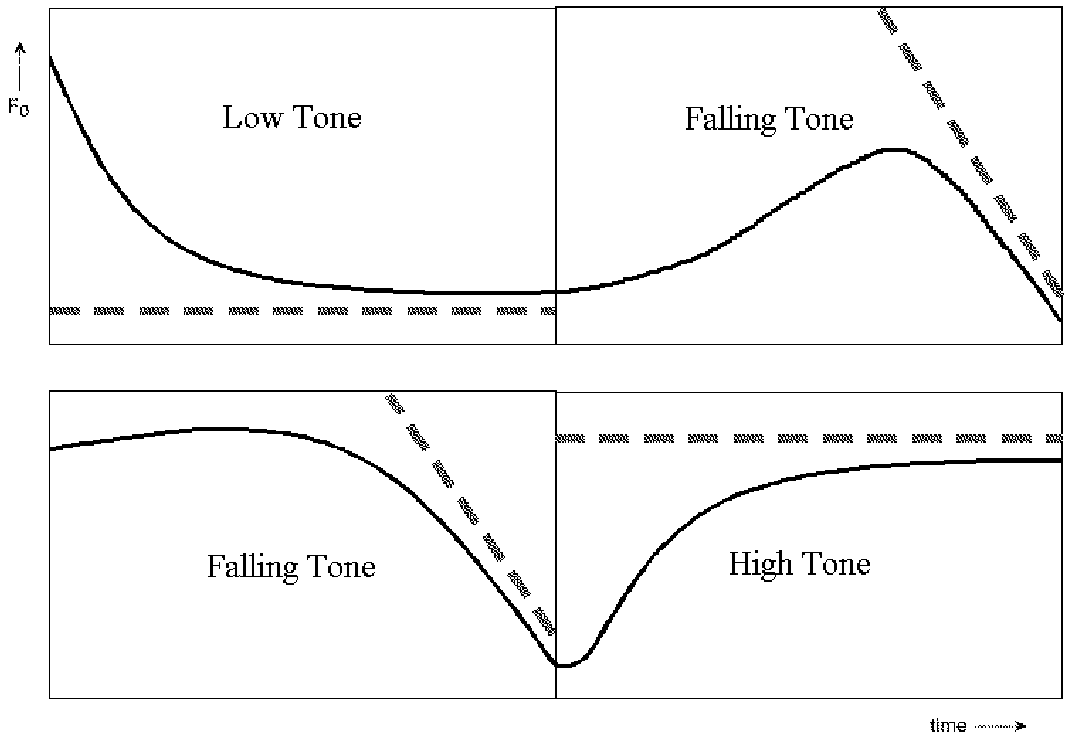

Figure 3. A schematic illustration of the F_0 contours of two successive tones based on the pitch target implementation model. The vertical lines represent syllable boundaries. The dashed lines represent underlying pitch targets. The solid lines represent the F_0 contours resulting from articulatory implementation of the underlying pitch targets.

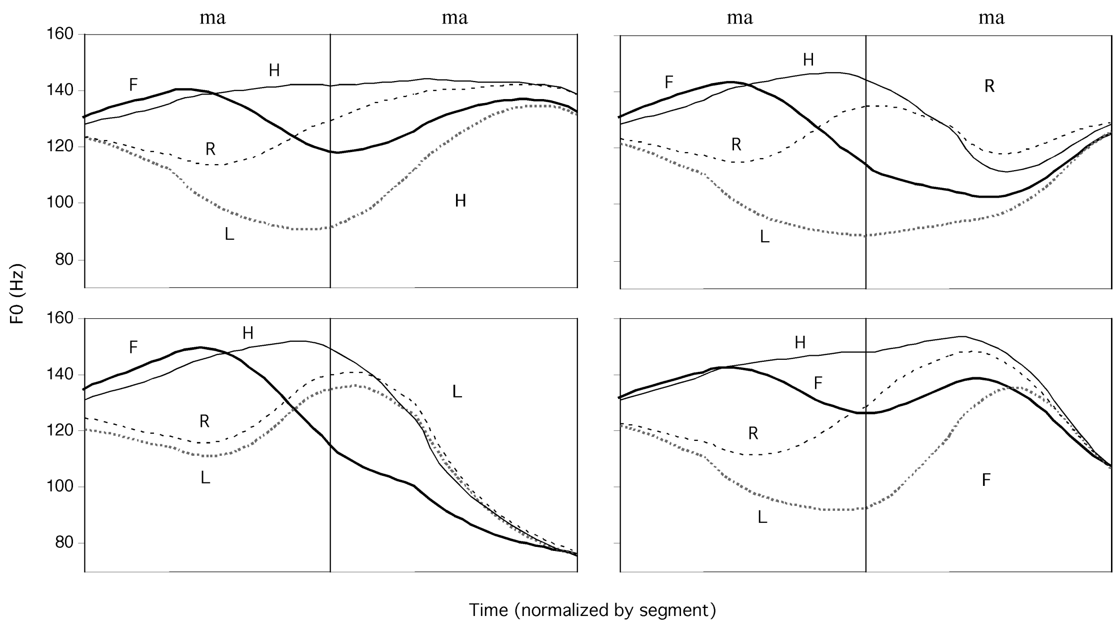

Figure 4. Effects of preceding tone on the F_0 contour of the following tone in Mandarin. In each panel, the tone of the second syllable is held constant, while the tone of the first syllable is either High, Rising, Low or Falling. The vertical lines represent the syllable boundaries (at the onsets of initial nasals). Each curve is a (segment-by-segment) time-normalized average of 192 tokens produced by eight speakers. (Adapted from Xu 1997)

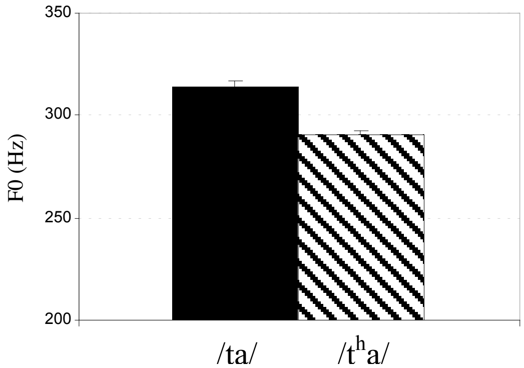

Figure 5. Average onset F_0 values of /ta/ versus /t^ha/.

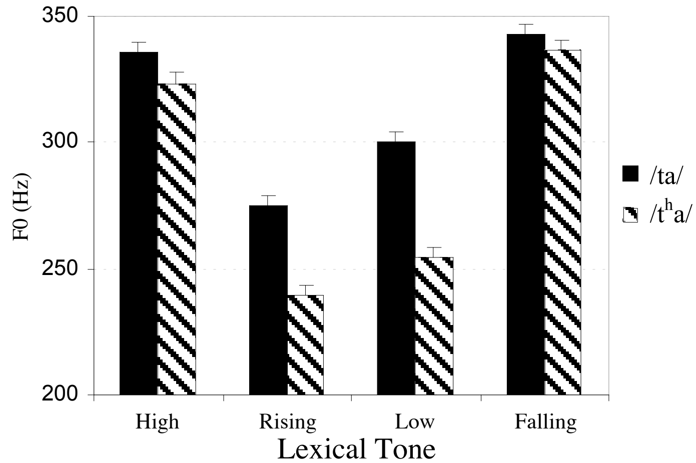

Figure 6. Average onset F_0 values of /ta/ versus /t^{h}a/ across four tones.

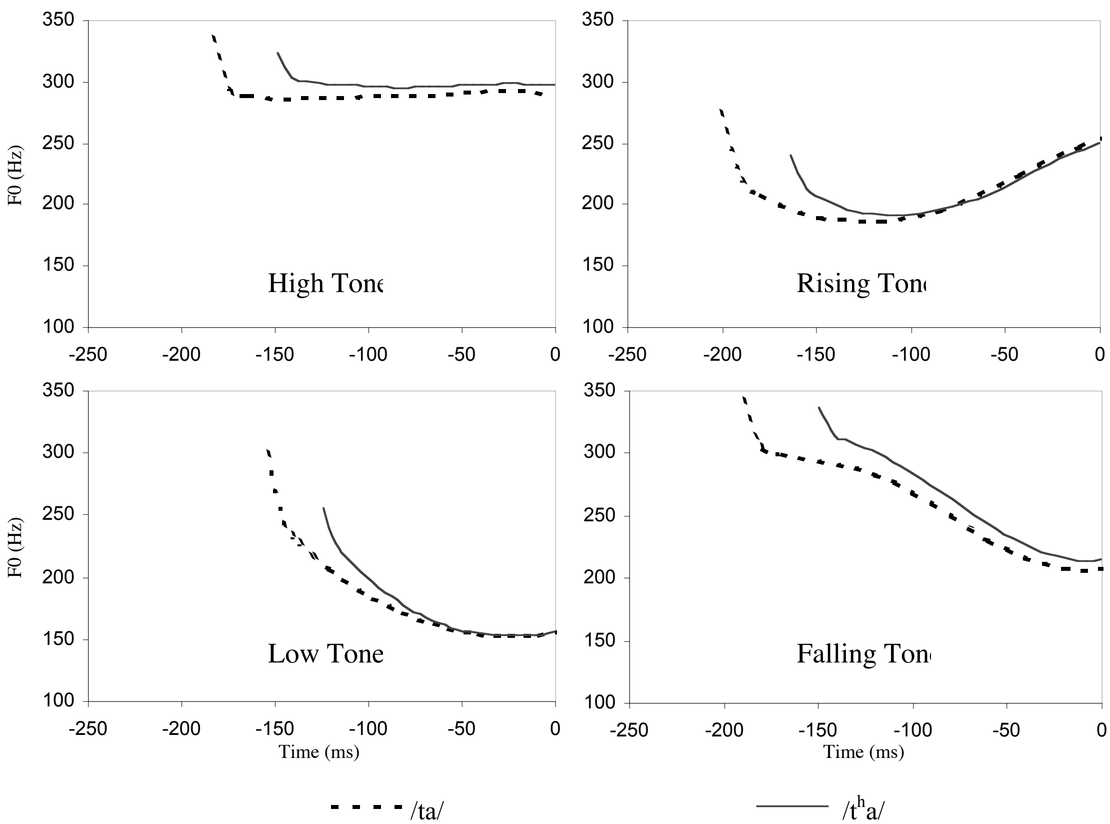

Figure 7. Average F_0 contours of syllable /ta/ versus /t^{h}a/ in four tones. The right boundary of each panel represents syllable offset.

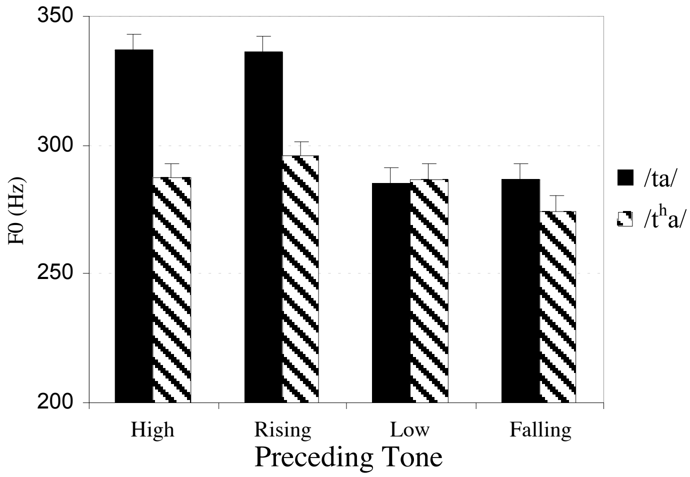

Figure 8. Average onset F_0 values of /ta/ versus /t^{h}a/ with different preceding tones.

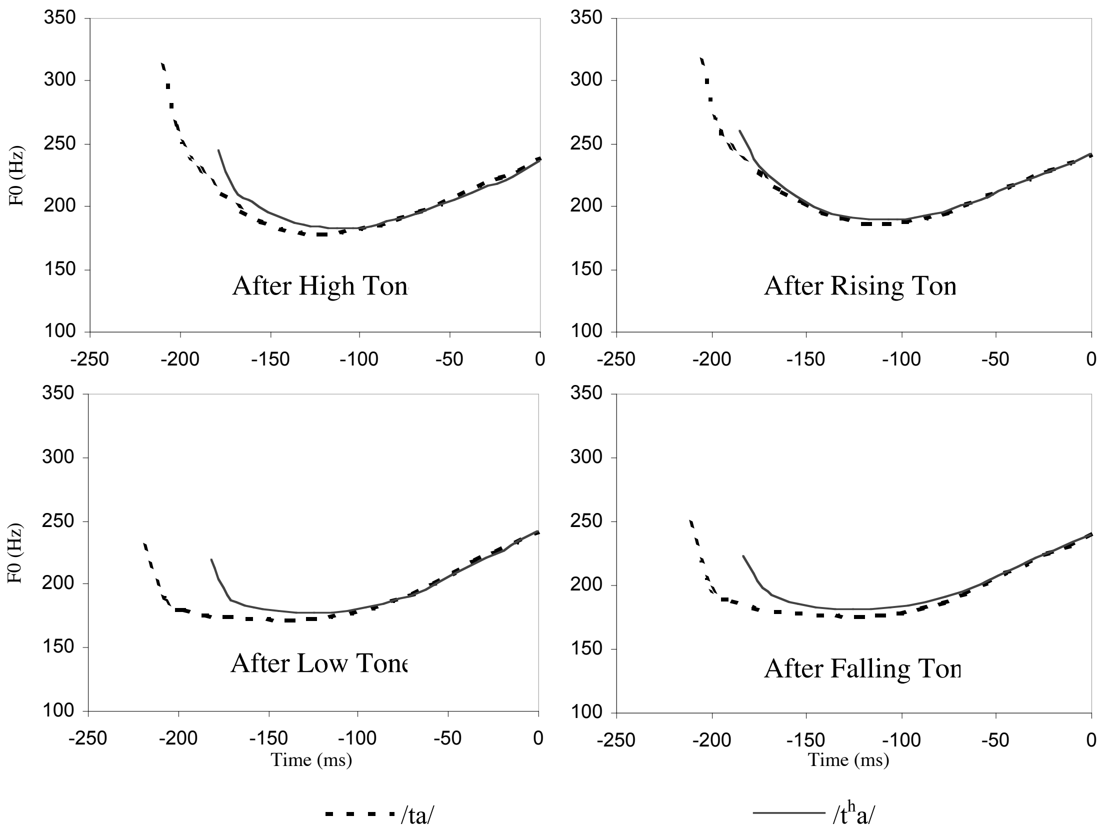

Figure 9. Average F_0 contours of /ta/ versus /t^{h}a/ with Rising Tone following different tonal contexts. The right boundary of each panel represents syllable offset.

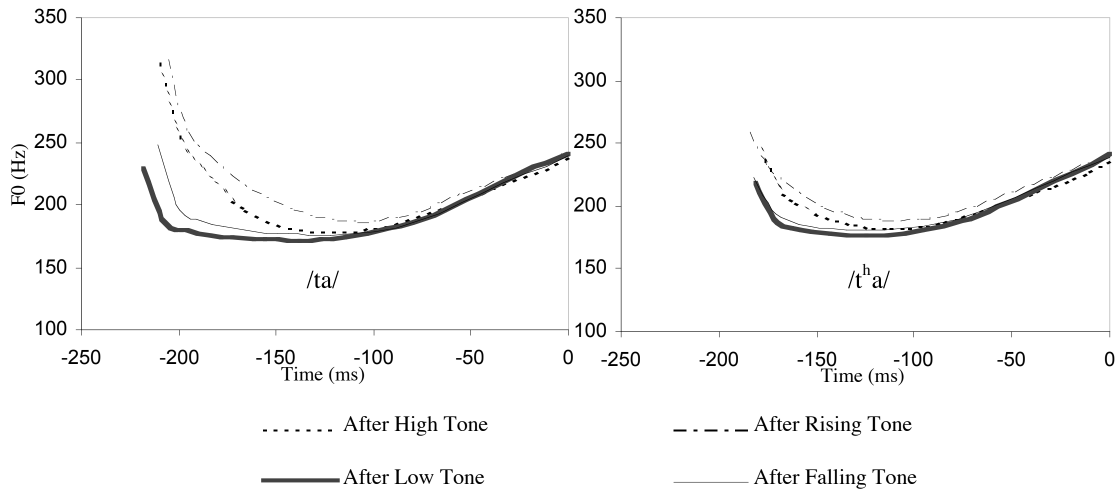

Figure 10. Average F_0 contours of /ta/ and /t^{h}a/ with Rising Tone following different tones. The right boundary of each panel represents syllable offset.

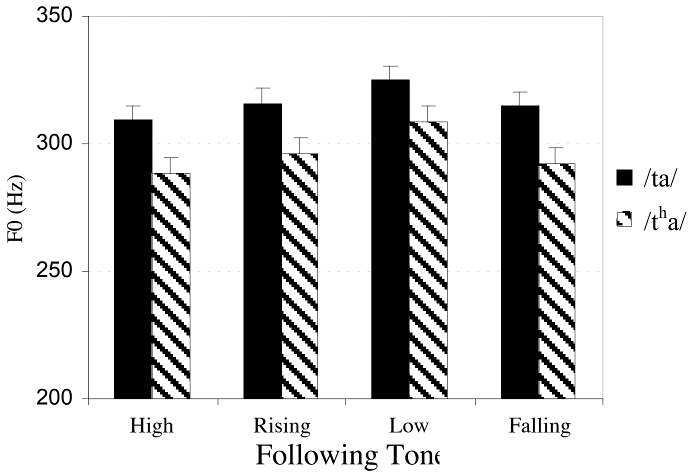

Figure 11. Average onset F_0 values of /ta/ versus /t^{h}a/ with different following tones.

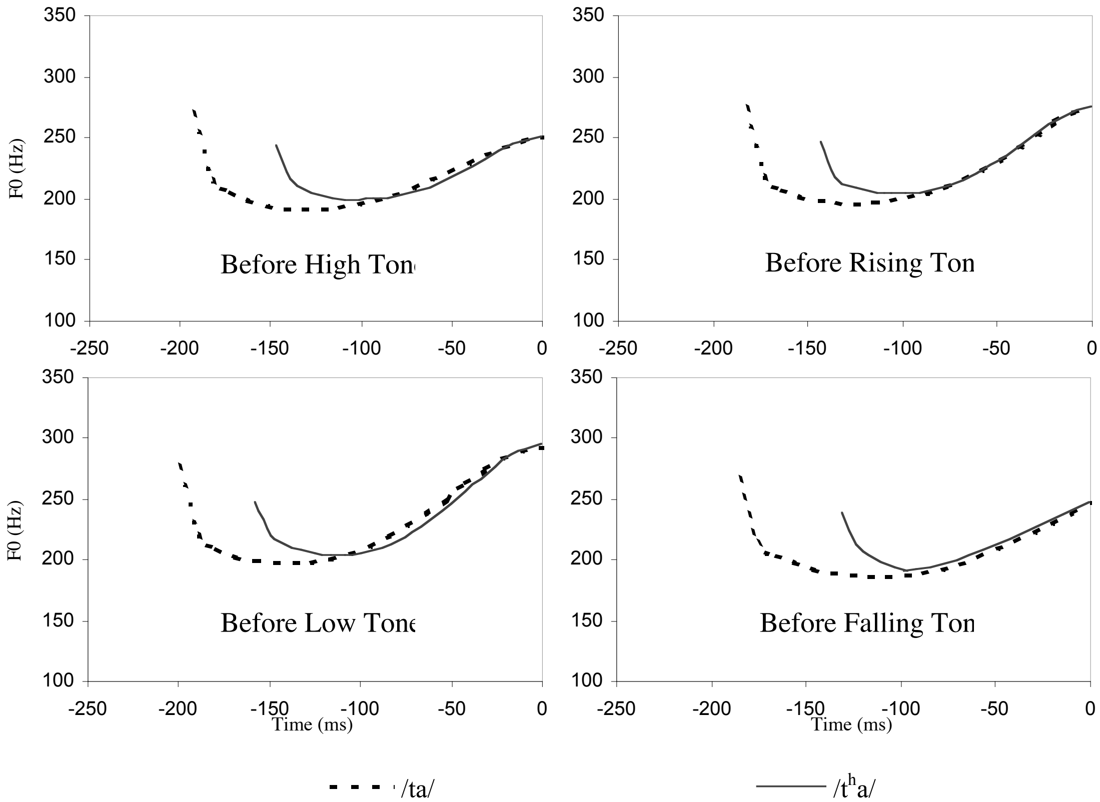

Figure 12. Average F_0 contours of /ta/ versus /t^{h}a/ with Rising Tone preceding different tonal contexts. The right boundary of each panel represents syllable offset.

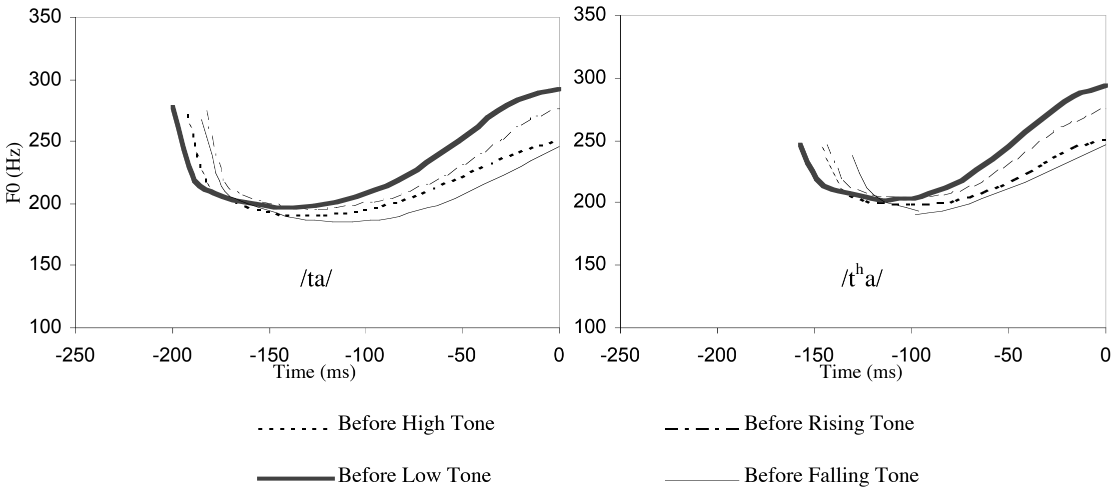

Figure 13. Average F_0 contours of /ta/ and /t^{h}a/ with Rising Tone when followed by different tones. The right boundary of each panel represents syllable offset.

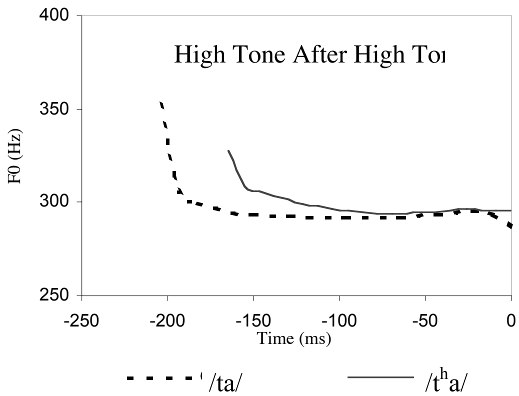

Figure 14. Average F_0 contours of /ta/ versus /t^ha/ of High Tone with High Tone preceding it.