Author version, Encyclopedia of Language and Linguistics, 2nd edition. Elsevier, Oxford. 12: 742-750.

Tone in Connected Discourse

Yi Xu

University College London

Haskins Laboratories

Dr. Yi Xu

Department of Phonetics and Linguistics

University College London

Wolfson House

4 Stephenson Way

London NW1 2HE

UK

yi@phonetics.ucl.ac.uk

Abstract:

Connected discourse has a number of substantial impacts on tone. First, the ideal form of tone is routinely compromised due to articulatory sluggishness. Second, the underlying tonal target itself is sometimes changed. Third, the surface tonal form is further distorted by other linguistic functions that also change pitch either deliberately or inadvertently. This article discusses how the surface acoustic form of tone can be understood in terms of the interaction between tone and these relatively independent processes.

In connected discourse lexical tones in a sentence have to be said in quick succession, frequently with no pauses in between. In addition to tone, a sentence also carries many other communicative functions that either use pitch as their primary or secondary carrier or change pitch inadvertently. As a result, the acoustic manifestation of tone, mainly in the form of fundamental frequency (F_0) contours, rarely resembles its underlying form. A clear understanding of tone is therefore impossible without comprehending the effects of these factors on the acoustic form of tone. This article discusses the realization of tone in connected discourse in terms of the interaction between tone and other pitch-modifying factors. The discussion is mainly based on data from Mandarin, for which extensive experimental data have been collected in recent years.

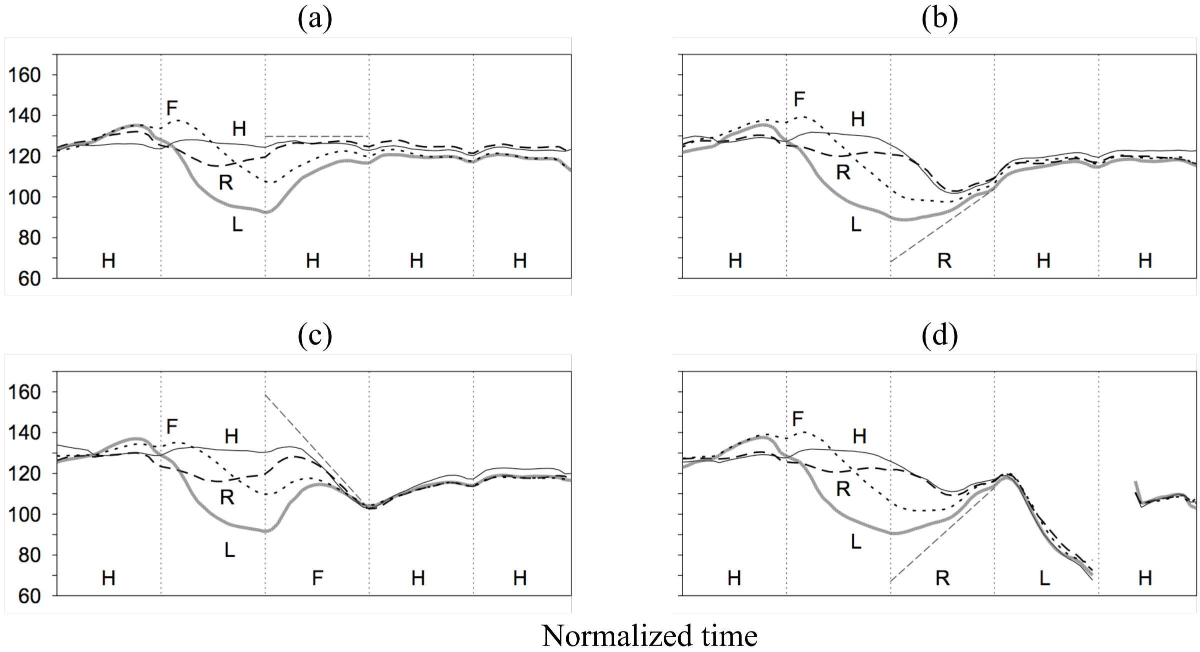

Figure 1. Mean F0 contours of Mandarin 5-tone sequences in sentence. Each curve is a time-normalized average across 24 repetitions by 4 male speakers (Xu 1999). (a)-(c): H, R and F (in syllable 3) preceded by four different tones and followed by H. (d): R in syllable 3 is followed by L. Vertical lines indicate syllable boundaries. The short dashed lines depict hypothetical underlying pitch targets. See supplementary slide 1 for audio samples.

HOW ARE TONES PRODUCED IN SUCCESSION?

In connected speech in a tone language, adjacent syllables frequently have different tones. One of the most basic issues about tone in discourse is then how a sequence of tones are articulated in succession. Is the shift from one tone to the next instantaneous, or fast enough so that the shifting period is negligible? If the shift is not instantaneous, where does it occur? Is there a time intervals specially reserved for the tonal shift? Answers to these questions can be found through systematic comparisons of sentences with different tonal compositions. Figure 1 displays time-normalized F_0 contours of Mandarin five-syllable sentences. In each panel the tones of all syllables are held constant except those of syllable 2, which alternate across High (H), Rising (R), Low (L) and Falling (F). Several direct observations can be made. First, although the tone of syllable 3 is the same in each panel, the F_0 trajectory there varies extensively when the tone of syllable 2 changes. Second, there is nevertheless consistency about the tones in syllable 3. In each panel, the four F_0 curves in the syllable gradually converge to a quasi-linear configuration appropriate for the tone: *high-level* for H, *rising* for R, and *falling* for F, as illustrated by the short straight dashed lines. Assuming that these quasi-linear trajectories resemble some sort of the desired goals for the tones, then the approximation of each of them starts roughly from the onset of syllable 3 and ends around the offset of syllable 3, whether the converging configuration is a slope or a horizontal line. These observations seem to suggest the basic mechanisms of tone implementation in connected speech. First, the transition between adjacent tones is not instantaneous. Rather, it takes a long time. Second, despite that, no time intervals are reserved just for the tonal transitions. Instead, the transition toward each tone occurs **in situ**, i.e., during the syllable to which the tone is designated. Third, each transition starts at the beginning of the host syllable and terminates at the end of the syllable. Thus the F_0 of an entire syllable constitutes a continuous movement toward the ideal pitch pattern of the tone it hosts. In other words, the approximation of the underlying pitch target of a tone is *synchronized* with its host syllable.

The sluggishness of the shift between adjacent tones is directly related to a basic articulatory constraint, i.e., the maximum speed of pitch change. This speed can be expressed with the following linear equations based on data reported in Xu and Sun (2002):

$$t = 89.6 + 8.7\,d \qquad \text{(raising)} \qquad [1]$$

$$t = 100.4 + 5.8\,d \qquad \text{(lowering)} \qquad [2]$$

where t is the amount of time (ms) it takes to complete the pitch shift, and d is the size of pitch shift in semitone (st). According to [1] and [2], it takes about 124 ms to either raise or lower pitch by 4 st. As can be seen in Figure 1, the difference between two adjacent tones is often close to or greater than 4 st (1 st = 12 * log2 (F_0 / $F_{0reference}$)). In connected speech, the duration of a syllable is very short when compared to the minimum time needed to make the tonal transitions. In Mandarin, for example, the average syllable duration is about 180-215 ms (Duanmu 1994; Xu 1999). This means that to make a single pitch shift of 4 st the greater half of a syllable has to be used just for the transition even at the maximum speed of pitch change. As can be seen in Figure 1, if the tone is R or F, often two pitch movements have to be made within one syllable. According to equations (1) and (2), there is barely enough time for *any* two pitch movements within a syllable of 180-215 ms long. Indeed, in producing these tones the maximum speed of pitch change is frequently approached (Xu & Sun 2002). In fact, even the maximum speed is often not fast enough. This is evident from the apparent "undershoot" in syllable 3 in Figure 1: when the tone of syllable 2 ends low, the F_0 peak in F becomes lower; and when the tone of syllable 2 ends high, the F_0 valley in R becomes higher. Furthermore, the location of these F_0 turning points also varies: the greater the F_0 excursion in the early part of the syllable, the later the turning point.

The tonal undershoot and the delay in the syllable-internal turning points further highlight the synchrony between tone and syllable. For without such synchrony, in the H L F sequence the transition toward F could have started earlier, given that L is more easily approached because only a single pitch movement is involved. Apparently, no micro-adjustment of tonal alignment is made and the synchrony between tone and syllable is maintained, often at the expense of full realization of the tone.

These facts suggest *syllable-synchronized sequential (pitch) target approximation* as the basic mechanism of tone implementation. A pitch target is an ideal pitch trajectory associated with a tone. A target can be rather simple, as illustrated by the short dashed lines in Figure 1, but still capable of leading to a complex surface F_0 contour due to the target approximation process. It can be a static register such as [high], [low] or [mid], which probably underlie a tone such as H, L or M. Targets can also be dynamic, such as [rise] and [fall]. For these targets, both the slope and relative height are part of the articulatory goal. Such targets may underlie tones like R and F.

A further implication of syllable-synchronized sequential target approximation under the constraint of the maximum speed of pitch change is that, if the duration of a syllable in a language is routinely shorter than that needed for two F_0 movements within a syllable, it may become impossible for that syllable to carry contour tones. Indeed, recent research has revealed a close relation between the occurrence of contour tones and average vowel length (which is closely related to syllable length) in many languages (Duanmu 1994; Zhang 2001 and references therein).

Finally, if the contextual F_0 variation of tone as shown in Figure 1 is viewed in terms of the influence of adjacent tones upon each other, there is extensive carryover assimilation but little anticipatory influence. If any, the anticipatory influence is mostly *dissimilatory* as can be seen in Figure 1: the F_0 of H in syllable 1 is higher whenever the tone of syllable 2 has a low F_0 point. Such dissimilation has been found in several languages, including Mandarin, Thai, Yoruba and Igbo. But its underlying mechanism is still unclear.

DYNAMIC VERSUS STATIC TARGETS

A long-standing discussion in tone research is over whether a tone such as R or F is composed of successive level components so that the underlying composition is actually LH or HL, or a single contour component that somewhat resembles the actual F_0 curve. The constraint of the maximum speed of pitch change makes it unlikely that the actual F_0 contour directly corresponds to an underlying configuration. On the other hand, whether or not two separate underlying components are involved can be tested in terms of whether a single target or two successive ones are being implemented. Figure 2 illustrates how such distinction can be made. If a particular tone consists of two elements, as indicated by the low and high level bars in Figure 2a, a clear transition should occur between the two. Thus a LH combination should result in an F_0 contour like the solid curve in Figure 2a, which approaches the [low] and [high] targets in succession. If, however, the target is unitary but dynamic, like the dashed slanting line in Figure 2a, the resulting contour should have the shape of the dashed curve. If the previous tone happens to end low, then the F_0 contours resulting from two static targets or one dynamic target may look like the solid or dashed curve in Figure 2b, respectively. A further test is that the distinction between consecutive static targets and unitary dynamic target should become even more apparent when (a) the syllable duration is increased, and (b) the tone is emphasized and hence given greater articulatory effort for implementing its target(s).

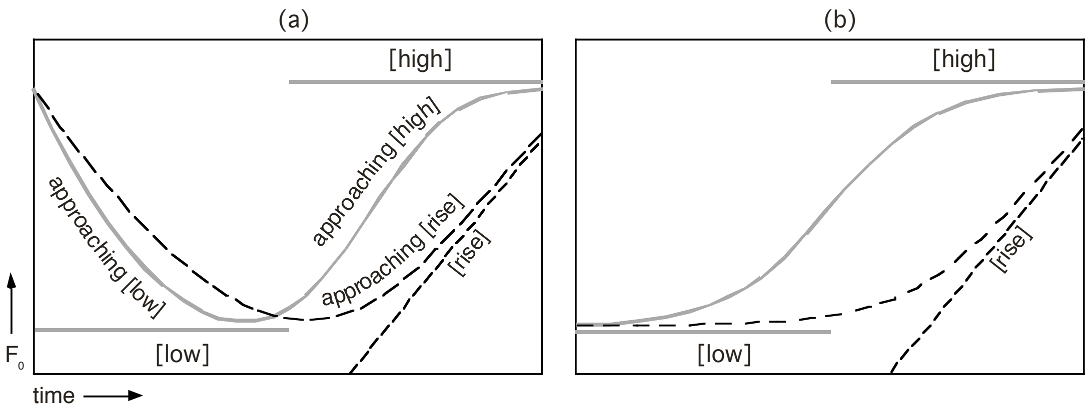

Figure 2. Hypothetical F_0 trajectories (curved lines) that asymptotically approach either two consecutive static targets (solid lines) or a single dynamic target (dashed lines). In (a) the previous tone ends high, while in (b) the previous tone ends low.

A case in point is R and F in Mandarin. When spoken at a slow rate, the F_0 contour of R resembles the dashed rather than the solid curves in either Figure 2a or Figure 2b, and the most dynamic portion of the F_0 contour becomes increasingly later as the syllable duration increases. Furthermore, when under focus (i.e., being emphasized), the dynamic portions of R and F in Mandarin occur even later in the syllable, as can be seen in Figure 3, instead of exhibiting clear separation of two static components as do the solid lines in Figure 2.

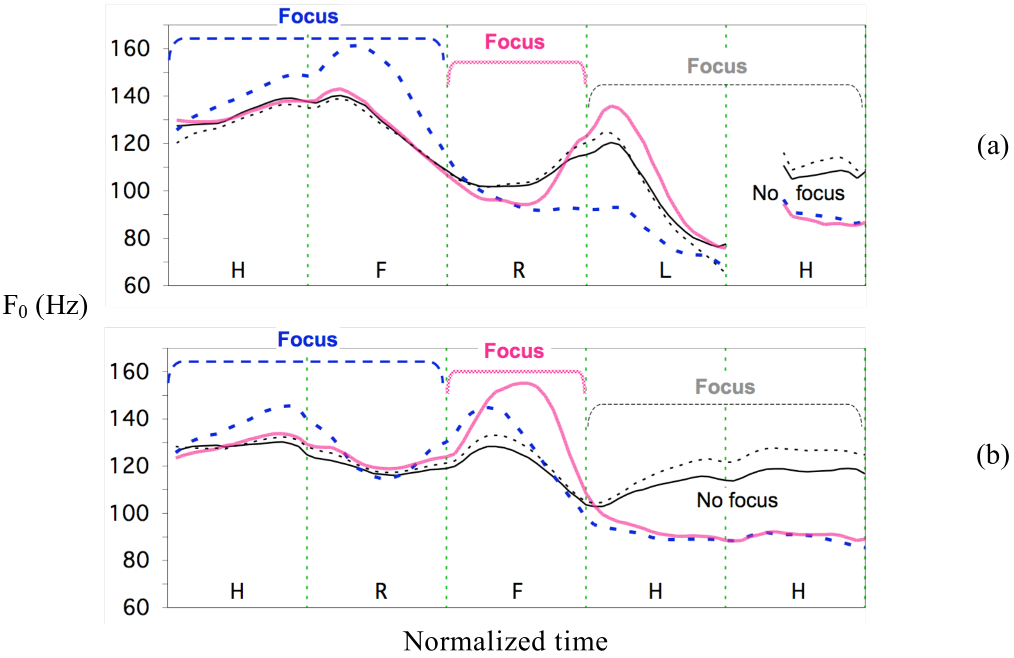

Figure 3. Mean F_0 contours of Mandarin tone sequences H F R L H (a) and H R F H H (b) spoken with focus in different positions. The curves are time-normalized F_0 averaged across 24 repetitions by 4 male speakers (data from Xu 1999). See supplementary file 2 for audio samples.

The evidence thus suggests that tones like R and F in Mandarin are associated with dynamic [rise] and [fall] as pitch targets. On the other hand, of course, the existence of dynamic tonal targets does not mean that any contour-like tone is underlyingly dynamic. It is also possible for a tone to be composed of two successive pitch targets. For example, the citation form of L in Beijing Mandarin may indeed consist of two components: a [low] followed by either a static [mid] or [mid-high] or a dynamic [rise]. (Further research is needed to determine which is more likely the case.) It is thus an empirical question whether a particular contour-like tone in a language or dialect has underlyingly static or dynamic target(s).

TONAL VARIATION DUE TO CHANGE OF UNDERLYING PITCH TARGET

The understanding of the basic mechanisms of producing tones in succession can also help us better identify cases where the targets themselves are changed. Such changes are often referred to as tone sandhi. The sandhi may be prompted by different triggers, some related to particular tonal contexts, some to morphological or syntactic functions (see Phonology of tone for more details). In the case of Mandarin L-sandhi, for example, L becomes very different when followed by another L than when followed by other tones, as can be seen in Figure 4a. Figure 4b shows that the surface form of such L variant is very similar in shape to that of R. Indeed, native listeners cannot distinguish L L from R L (Wang & Li 1967; Peng 2000).

What is most striking about this kind of tonal variation is that, as can be seen in Figure 4a, the sandhi L starts to differ very early in its host syllable from the non-sandhi L, and the latter exhibits virtually no sign of anticipating the following tone. This suggests that what happens to L in anticipation of an upcoming L is that its pitch target is *changed before execution*, and such change does not alter the process of syllable-synchronized sequential target approximation. In other words, once assigned, a sequence of pitch targets are just implemented one at a time: each during the time interval of its host syllable. No anticipatory target execution is involved.

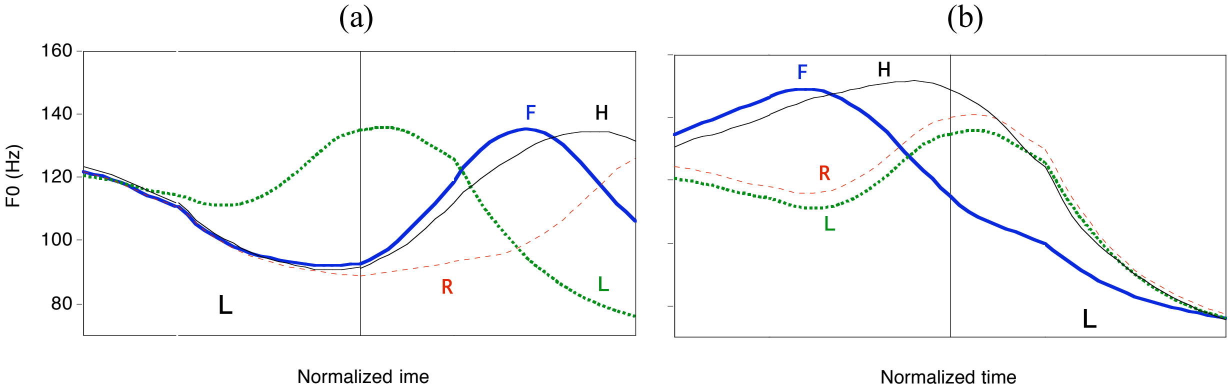

Figure 4. (a) Mean F0 contours of Mandarin L before (a) and after (b) four different tones. (Adapted from Xu 1997). See supplementary file 3 for audio samples.

Target alternation is not limited to anticipation of an upcoming tone, it can also happen *after* certain tones. In Yoruba, for example, H is known to become R after L, and L become F after H. It is not yet fully clear, however, whether those are genuine cases of target alternation or rather cases of carryover assimilation. Further research is needed.

TONAL VARIATION RELATED TO ASPECTS OF TARGET EXECUTION

While tones specify the pitch targets (which are subject to alternation as just discussed) of individual syllables, other aspects of the target approximation process are left unspecified, and hence available for manipulation to encode other communicative functions. These aspects include articulatory strength, syllable duration, and pitch range. The manipulation of these aspects has strong effects on the surface F_0 contours. In fact, F_0 variations due to some of these effects are often no smaller than those due to the lexical tones themselves.

Tonal variation due to articulatory strength and syllable duration

Articulatory strength determines how quickly a pitch target is approximated, and syllable duration determines how much time is allotted to the approximation process. Other things being equal, therefore, greater strength enables a pitch target to be approached more quickly than weaker strength, and longer syllable duration allows a target to be more fully implemented than shorter duration. Thus manipulation of either strength or duration or both determines the extent to which a lexical tone is realized. And such manipulation can be used to encode certain communicative information.

Strength and duration as lexical information

In some tone languages, certain morphemes are considered to be phonologically toneless (see the Phonology of Tone), e.g., those morphemes in Mandarin that are said to bear the neutral tone. Also L in Chichewa and M in Yoruba are believed to be unspecified for tonal values. As to where the F_0 values of these syllables come from, a popular view is that they are derived through interpolation between adjacent tones. Recent experimental data suggest, however, that interpolation is an unlikely process. Figure 5 shows Mandarin sentences containing three consecutive neutral tone (N) syllables. In (a) they are preceded by four different tones and followed by F. In (b) they are preceded by H but followed by either F or L. In (a) the F_0 of the neutral tone syllables is substantially affected by the preceding tone. But the influence gradually reduces as time elapses, and by the end of the third neutral tone, the four F_0 curves have almost converged. In (b), in contrast, the tone of the last syllable has virtually no influence on the preceding neutral tone. Thus F_0 of the neutral tone could not have been derived from interpolation between the surrounding tones, because the following tone makes no contribution to its F_0. Given that the carryover influence from the preceding tone is similar in nature as that seen in Figure 1, the F_0 convergence over the course of the neutral tone must be in the direction of an underlying pitch target associated with the neutral tone itself, which is probably a static [mid]. What makes the neutral tone really different from the full tones is likely to be weaker articulatory strength plus shorter syllable duration (about 60% of the length of full tone syllables), which would lead to both slower target approximation and greater undershoot.

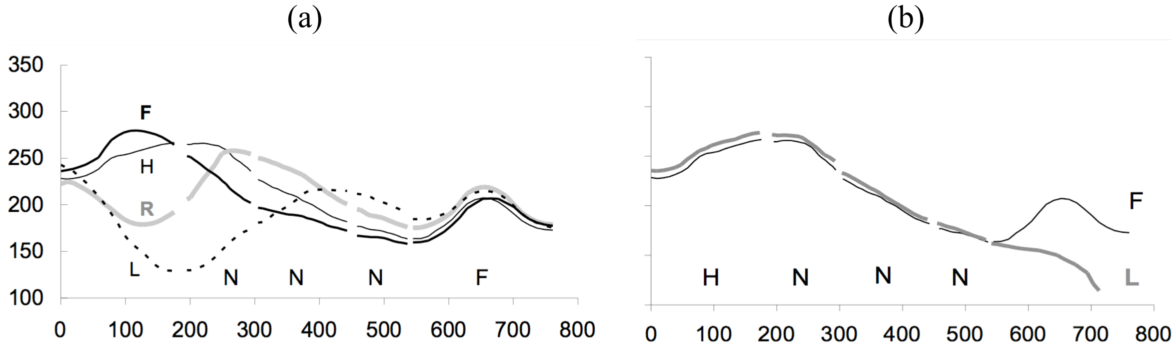

Figure 5. Mandarin sentences containing three consecutive neutral tone (N) syllables. In (a) they are preceded by four different tones. In (b) they are followed by either F or L. Data from Chen and Xu (2002). See supplementary file 4 for audio samples.

As to how weak the strength can be, there have been suggestions that the level of articulatory effort is zero, and the F_0 movement during the neutral tone is due to relaxation to a rest position under the elastic force of the vocal folds. Note that such elastic force is available only if the vocal folds are stretched beyond their resting length during the production of the full tones. However, the vocal folds are typically much shorter during phonation than during rest (Hollien & Moore 1960). Thus whatever pitch target is associated with a neutral tone or its like, it is probably implemented with active muscle forces, although the forces can be quite weak at times.

Strength and duration as grouping and demarcation information

In addition to lexical information, strength may also serve to group syllables into chunks, and to group smaller chunks into larger ones. Such grouping is closely related to a hierarchical prosodic structure whose communicative function is not yet fully clear. What is determined by prosodic structure is the grouping relation among the linear sequence of syllables and words. It has been demonstrated that a syllable’s ability to resist tonal coarticulation from adjacent tones is related to its prosodic strength (Shih & Sproat 1992). Tones on prosodically weak syllables tend to have less extreme tone shapes, or, greater undershoot.

Figure 6 displays mean F_0 contours of consecutive R and F sequences with varying numbers of syllables said in the middle of a carrier. The all-R and all-F sequences impose great pressure on tone production, as discussed earlier, and hence would best reveal the strength differences among syllables. As can be seen in Figure 6, as the number of syllables increases, the magnitude of F_0 excursion decreases, which appears to be related to the insufficient increase in syllable duration as the group size becomes larger. Meanwhile, the magnitude reduction differs in various locations. But in each case it is the first and last syllable in the group that are reduced the least. There thus appears to be an “edge marking” effect. This line of research is still in its early stage, and much more needs to be learned about how exactly articulatory strength is used by various functions.

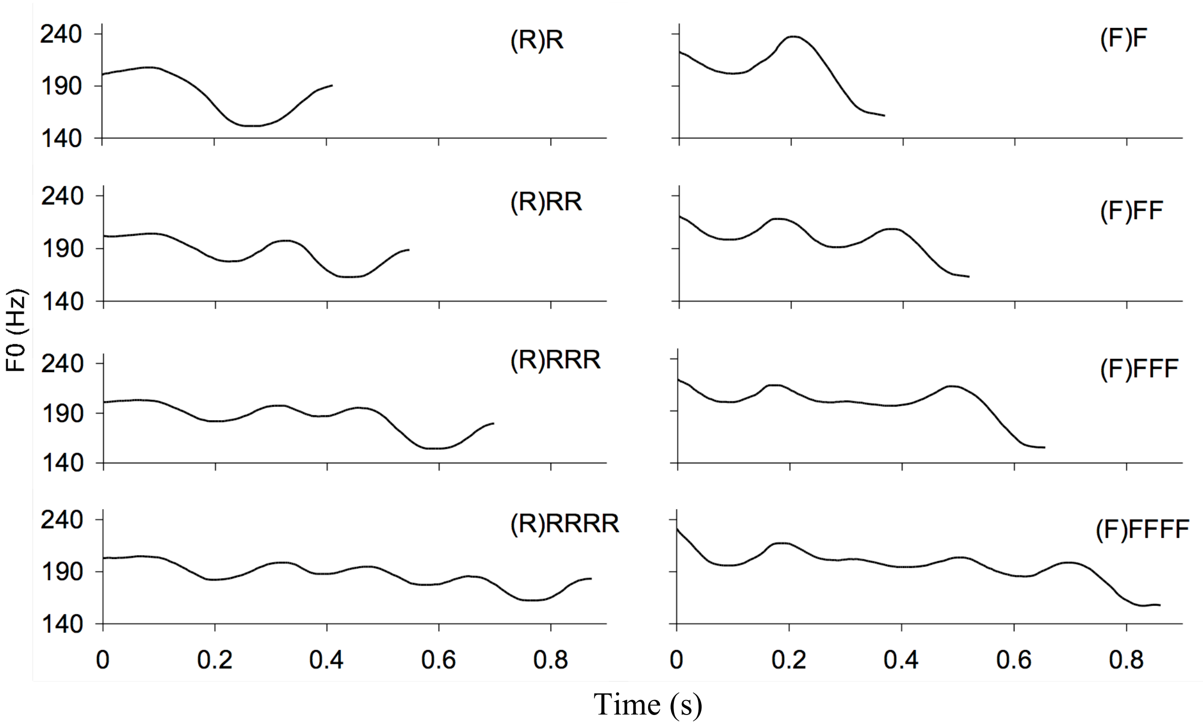

Figure 6. Mean F_0 contours of consecutive R and F sequences with varying numbers of syllables said in middle of a carrier. See supplementary file 5 for audio samples.

Tonal variation due to pitch range modification

Pitch range specifies the pitch interval within which each tonal pitch target is implemented. It is definable by two parameters: height and span. The effectiveness of pitch range manipulation has to do with the fact that the entire exploitable pitch range of human speakers is quite large, about 2 octaves in total (Fairbanks 1959). This is more than twice as much as the range needed for lexical tones. Thus even in a tone language there is room left in the pitch range for things other than lexical tones. From the perspective of the target approximation process, pitch range manipulation can be achieved by shifting the height of a pitch target without changing anything else about it. Pitch range manipulation has been put to use by languages for encoding functions such as focus, sentence type and new topic (topic shift/turn taking).

Focus

Focus is a pragmatically motivated emphasis. For example, when the sentence "John talked Jane" is said in response to the question "Who talked to Jane?" the word "John" is naturally emphasized, hence, "focused." Under certain circumstances, speech sounds as small as a single segment can be put under focus (van Heuven 1994). At the same time, focus is a global function. That is, with focus the speaker tries to indicate that a particular component of the utterance is highlighted against all other components. This is done by an asymmetrical pitch range manipulation about the focused item, as can be seen in Figure 7 as well as Figure 3 shown earlier. Specifically, a narrow focus is manifested through a tri-zone pitch range control: the pitch range of focused region is expanded; that of the post-focus region suppressed; and that of the pre-focused region left largely the same as in utterances without narrow focus. Expansion means the pitch range is widened instead of only raised. The widening is the most obvious in Figure 3a where the low point of R becomes lower under focus. Tri-zone pitch range control by focus is quite widespread among languages, both tonal and non-tonal, although it is not yet clear whether it is truly universal.

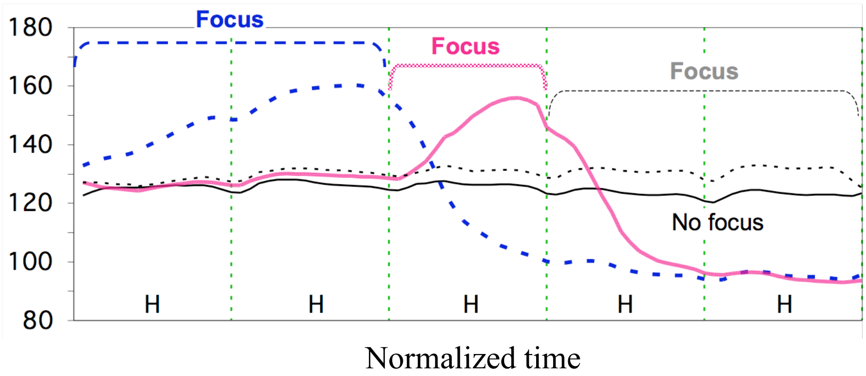

Figure 7. Pitch range variations in Mandarin due to focus in different positions. The curves are time-normalized F_0 averaged across 24 repetitions by 4 speakers (data from Xu 1999). See supplementary file 6 for audio samples.

An interesting consequence of the tri-zone pitch range control is that a final focus, i.e., focus on the last word of a sentence, is not implemented effectively. This is clearly seen in both Figure 3 and Figure 7. As a result, final focus is also perceptually not well separated from no narrow focus. Presumably, the on-focus pitch range expansion is conditioned by post-focus pitch range suppression: if the latter is not implementable (for there is no post-focus component), the former cannot be effectively implemented either.

Sentence type: statement vs. question

Pitch range can be further modified by a function that may be referred to as sentence type. This function distinguishes question from statement, and possibly also from continuation. A question is known to have an overall rising intonation, although it is also known that not all syntactic questions have a rising intonation. Although the F_0 difference between question and statement is the most obvious at the end of the sentence, more global patterns are also involved. In particular, the onset of the F_0-raising by question is conditioned by focus. Figure 8 shows mean F_0 contours of Mandarin question versus statement in sentences consisting of only H. In Figure 8a focus is either on the sentence initial word, or there is no narrow focus (neutral focus). In Figure 8b, focus is either sentence medial or sentence final. It is apparent that the divergence between statement and question starts from the focused word rather than from the beginning of the sentence, or occurring only in the final word of the sentence.

The F_0 patterns in Figure 8 thus demonstrate that question intonation consists mainly of pitch range adjustments, and the adjustments are likely to be nonlinear over time, with greater increase toward the end of the question. The nonlinearity may explain why the final raising is typically much more prominent than in earlier regions. Similar to the nonlinear F_0 raising in questions, the gradual F_0 lowering in statements is also likely nonlinear, with greater decrease toward the end of the statement.

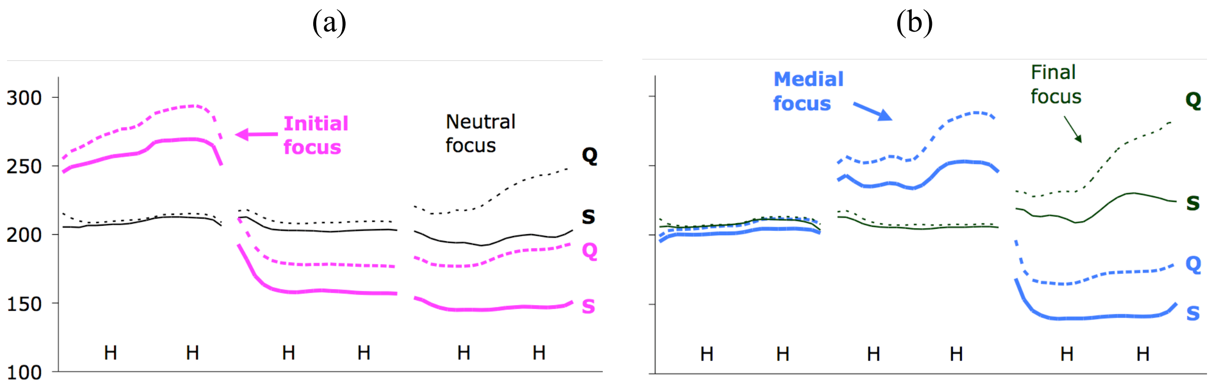

Figure 8. Mean F_0 contours of Mandarin question versus statement in sentences consisting of only H tones. In (a) Either focus is on the sentence initial word, or there is no narrow focus. In (b) Focus is either sentence medial or sentence final. Data from Liu and Xu (2004). See supplementary file 7 for audio samples.

New topic (topic shift / turn taking)

Pitch range can be further modified by a function that affects the beginning of a sentence more than the end of the sentence. When a sentence is the very first in a conversational turn or a read paragraph, its onset F_0 can be extensively raised, sometimes by as much as an octave (Umeda 1982). While the details are still not quite clear, this function also seems to affect the pitch range of a sentence non-linearly, with a large F_0 raise near the beginning of the sentence and an F_0 descent afterwards that is fast at first but slowing down later on. The high initial F_0, often accompanied by increased intensity, probably helps to draw listeners' attention. As such, the function probably serves as a beginning signal for a new topic or a conversational turn.

The effects of pitch range manipulations combined

We have seen that although the local F_0 contours of individual syllables are mostly determined by the underlying pitch targets and their manner of implementation as related to articulatory strength and syllable duration, the relative height of the tones can still be extensively modified by focus, sentence type and new topic. Note also that these functions, with the only exception of question intonation, all tilt F_0 downward toward the end of the sentence. Such tilting is in the same direction as the frequently reported tonal phenomenon known as downdrift. It thus awaits future research to determine if downdrift is in any way related to these functions.

INADVERTENT F_0 VARIATION

Inadvertent F_0 variation refers to F_0 changes brought by functions whose primary target of control is not F_0. Two types of inadvertent F_0 variations have been widely reported: vowel intrinsic F_0 and perturbation of F_0 by consonant. As their names indicate, the primary target of control is the articulation of the vowel or consonant rather than F_0 per se. Nevertheless, F_0 is inadvertently affected by the articulatory manoeuvres involved.

Intrinsic F_0 of vowel

Intrinsic vowel F_0 refers to the phenomenon that, other things being equal, different vowels are produced with different F_0 values. Such variation is mostly related to vowel height: the higher the vowel, the higher the F_0. The intrinsic F_0 difference is in the range of 1.65 semitones or 15 Hz (Whalen & Levitt 1995). Intrinsic F_0 has been found in both tone and non-tone languages. Being small in magnitude, it does not usually affect either the local or global F_0 patterns in a significant way. There is also evidence that its magnitude becomes smaller in connected speech (Ladd & Silverman 1984). Nevertheless, when it comes to systematic analysis of any effect on F_0 by other factors, intrinsic F_0 has to be taken into consideration.

Consonant perturbation of F_0

A non-sonorant consonant perturbs F_0 in two ways. First, it interrupts the otherwise continuous F_0 flow. Second, it either raises or lowers the F_0 of the adjacent vowels. While the F_0 raising and lowering have been widely known, the effect of F_0 interruption is often not carefully considered. In general, it has been taken for granted that because of the F_0 interruption, consonants are not part of the tone-bearing unit. Recent research, however, has found evidence that the voicing interruption does not really disrupt the pitch target approximation process as seen in Figure 1. Figure 9 displays the mean F_0 contours of Mandarin syllables /ma/, /da/, /ta/ and /sha/ spoken with R and F. Compared to the F_0 contours in /ma/, in which the transition toward the tonal target is visible, the F_0 curves in syllables with the voiceless consonants start late and have various amounts of local perturbations at the voice onset. Nonetheless, if these local effects are put aside, the F_0 curves in /da/, /ta/ and /sha/ look very similar to those of /ma/. Hence, by the time the apparent local effect are over, F_0 is already quite low in R but quite high in F. So, the process of syllable-synchronized sequential target approximation remains the same whether or not voicing continues through the consonant. Thus the effect of voiceless consonants is only to introduce rather local perturbations without changing the basic mechanism of tone production.

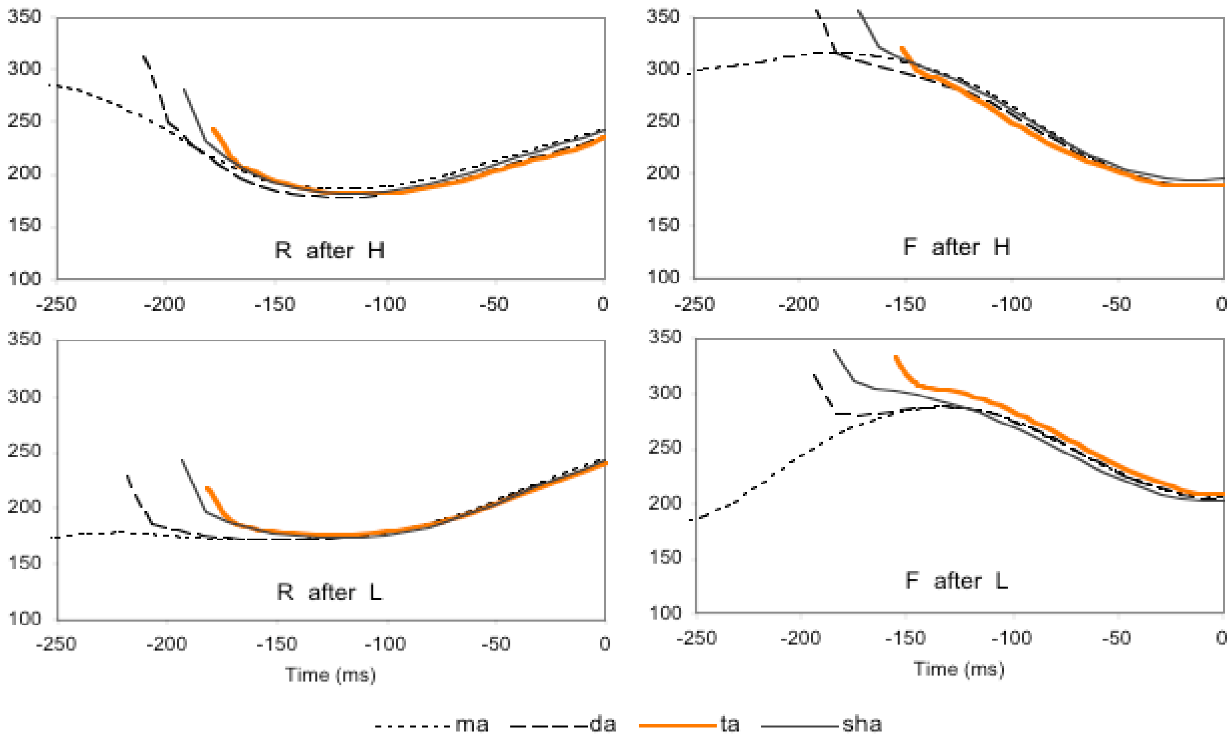

Figure 9. Effects of voiceless consonants on the F_0 contours of Mandarin R and F produced after H and L. Each curve is an average across 5 repetitions, 2 carrier sentences and 7 female speakers. All curves are aligned to the syllable offset. Adapted from Xu et al. (2003).

See also:

Neurophonetics of Tone; Phonology of Tone; Intonation; Intonation and syntax; Prosodic cues of discourse units; Prosodic Aspects of Speech and Language; the phonology of (Mandarin) Chinese; Coarticulation; Speech Anatomy; Speech Production; Syllable; Phonology-Phonetics interface; Speech Synthesis, Prosody; Emphasis.

FURTHER READING:

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

Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge, UK: Cambridge University Press.

Chen, Y. and Xu, Y. (2002). Pitch Target of Mandarin Neutral Tone. LabPhon 8, New Haven, CT. New Haven, CT.

Duanmu, S. (1994). Syllabic weight and syllable durations: A correlation between phonology and phonetics. *Phonology* 11, 1-24.

Fairbanks, G. (1959). *Voice and Articulation Drillbook*. New York: Harper & Row.

Gandour, J., Potisuk, S. and Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics* 22, 477-492.

Hollien, H. and Moore, G. P. (1960). Measurements of the vocal folds during changes in pitch. *Journal of Speech and Hearing Research* 3, 157-165.

Kochanski, G., Shih, C. and Jing, H. (2003). Quantitative measurement of prosodic strength in Mandarin. *Speech Communication* 41: 625-645.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Ladd, D. R. and Silverman, K. E. A. (1984). Vowel intrinsic pitch in connected speech. *Phonetica* 41: 31-40.

Laniran, Y. O. and Clements, G. N. (2003). Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics* 31, 203–250.

Liu, F. and Xu, Y. (2004). Asking questions with focus. *Journal of the Acoustical Society of America* 115, Pt. 2: 2397.

Myers, S. (1999) Surface underspecification of tone in Chichewa. *Phonology* 15.3, 367-392

Peng, S.-h. (2000). Lexical versus 'phonological' representations of Mandarin Sandhi tones, in: M. B. Broe and J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge University Press, Cambridge: 152-167.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT, Cambridge, MA. [Published in 1987 by Indiana University Linguistics Club, Bloomington.]

Shih, C. and Sproat, R. (1992). Variations of the Mandarin rising tone. *Proceedings of The IRCS Workshop on Prosody in Natural Speech* No. 92-37, Philadelphia. The Institute for Research in Cognitive Science, University of Pennsylvania: 193-200.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7, 71-79.

Swerts, M. (1997). Prosodic features at discourse boundaries of different length. *Journal of the Acoustical Society of America* 101, 514-521.

Thorsen, N. G. (1980). A study of the perception of sentence intonation — Evidence from Danish. *Journal of the Acoustical Society of America* 67, 1014-1030.

Umeda, N. (1982). "F0 declination" is situation dependent. *Journal of Phonetics* 10, 279-290.

van Heuven, V. J. (1994). What is the smallest prosodic domain? in: P. A. Keating (Ed.). *Papers in Laboratory Phonology*. Cambridge University Press. 3: 76-98.

Wang, W. S.-Y. and Li, K.-P. (1967). Tone 3 in Pekinese. *Journal of speech and Hearing research* 10, 629-636.

Whalen, D. H. and Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics* **23**: 349-366.

Xu, C. X., Xu, Y. and Sun, X. (2003). Effects of Consonant Aspiration on Mandarin Tones. *Journal of the International Phonetic Association* 33, 165-181.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27, 55-105.

Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics* 5, 757-797.

Xu, Y. and Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111, 1399-1413.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Zhang, J. (2002). *The Effects of Duration and Sonority on Contour Tone Distribution -- A Typological Survey and Formal Analysis*. New York: Horn, Laurence.