The phonetics and phonology of apparent cases of iterative tonal change in Standard Chinese

Yu-ching Kuo, Yi Xu and Moira Yip
University College London

Final version

All correspondence to:
Moira Yip
Dept of Phonetics and Linguistics
University College London
Gower Street
London WC1E 6BT
UNITED KINGDOM

moira@ling.ucl.ac.uk

The phonetics and phonology of apparent cases of iterative tonal change in Standard Chinese

Yu-ching Kuo, Yi Xu and Moira Yip
University College London

Standard Chinese has four lexical tones: high (H), low (L), rising (R) and falling (F).[i] There are also the so-called 'neutral-toned' syllables, often viewed as toneless, although see Chen and Xu 2004 for a dissenting view. The language has a well-studied rule that changes the first of two low tones to a higher rising tone. The rule is known as the third-tone sandhi rule. It applies whenever the two low tones are adjacent in a domain, where the domains are defined by a complex mixture of syntactic, prosodic, and focus-related factors. It is partially dependent on speech rate, so that a given utterance may have more than one way of applying the sandhi rule. In utterances with several low tones in a row, the rule may apply to some or all of them, depending on the factors just mentioned. For example, Chen (2000:386) gives two different pronunciations for this phrase:

(1) /LLLH/ → RLLH *or* RRLH *[xiang [xie [xiao shuo]]]*
plan write novel
'to plan to write a novel'

If the syntactic structure is left-branching, the rule may apply several times to give a sequence of high rises, as in this example:

(2) /LLLLL/ → RRRRL *[[[[zong tong] fu] li] you]*
president's palace inside there-is
'inside the president's palace there is…'

This is usually attributed to cyclic rule application (see Chen for details), but it has also been suggested that the rule applies either from left-to-right, or simultaneously, to all the lows at once. Indeed, in sufficiently fast speech multiple low-to-rising (L-to-R) sweeps are possible even if the syntactic structure is right-branching, as the RRLH pronunciation of (1) shows. In natural speech, long sequences of rises like this seem to be rare, partly because in fast speech the rises then flatten out into a high plateau, an effect usually attributed to a second sandhi rule, which converts rises (whether underlying or not) to high level tones after a tone that ends high.

Although these facts have received extensive attention in the literature, some aspects remain little understood. In this paper we focus on two rather separate issues: what exactly happens in long sequences of low tones, and what is nature of the interaction between sandhi and prominence. The standard description raises several issues:

1. Given that the underlying lows often surface as flattened out to high level, can we find evidence for the intermediate stage at which they are all rising?
2. Can all cases where several lows become rising be attributed to cyclicity?

3. If not, how can such apparently iterative rule application be modelled in a non-derivational grammar like Optimality Theory?

In understanding the sandhi process, it will also turn out to be necessary to investigate how the sandhi rule interacts with focus and prosody in general.

This paper will propose answers to these questions, and through them shed some light on the relationship between phonology and phonetics when tones are produced in combination. In section 1 we briefly summarize the previous literature. Sections 2 and 3 detail two experiments that search for cases where strings of lows become rises in the absence of cyclic structure (L-to-R sweeps), and also investigate the effects of speech rate on the flattening of rises to high level pitch. We conclude that L-to-R sweeps are indeed found, and must be phonologically explained, but that the levelling of rises to high plateaux is gradient and phonetic, not phonological. In this section we also argue that underlying rises and derived rises have different phonetic targets, and are thus phonologically distinct. Section 4 begins with some background on tone-prominence interaction, then proposes a one-step Optimality Theory analysis for the facts, showing that there is no need to posit an iterative and thus derivational process. The analysis depends on allowing the Obligatory Contour Principle (OCP) to apply at all levels in the tonal structure, combined with an avoidance of solitary low tones on prosodic heads.

1. Previous descriptions of strings of L tones

As we have mentioned, strings of L tones may under certain circumstances change into strings of R tones. We shall refer to these as L-to-R sweeps. They are reported extensively in the phonological literature (Cheng 1973, Shih 1986, 1997, Chen 2000), where it is claimed that they (i) appear at all speech rates in left-branching structures, and also (ii) in all structures in fast speech. However, they seem to be quite rare in natural speech. Out of 54 utterances (statements and questions) containing long strings of up to nine L tones in Shen (1990), only two had RR sequences, and none had more than two R in a row.

A third context, usually not mentioned, is found in unstructured strings such as strings of digits, or foreign place names. Here we find such sweeps even in slow or normal speech, and they cannot be attributed to cyclicity.

These observations suggest that there are three sources for surface L-to-R sweeps:

(3) Cyclically in left-branching structures:
First cycle: [L L] L → [R L] L
Second cycle: [R L L] → [R R L]

(4) Left-to-right (or simultaneously) in slow speech ternary feet, as in odd-parity strings of digits:
'five-five-five'[L L L] → [R L L]→ [R R L]

(5) Left-to-right (or simultaneously) in fast speech, presumably as a result of domain enlargement:

[L L L] → [R L L] → [R R L]

Any theory which has some mechanism akin to cyclicity can explain the data in (1), but the other cases are potentially problematic. In an output-based theory, there are no derivations, and thus no intermediate stages (except perhaps from cyclic operations). Thus, left-to-right iterative sweeps, whereby /LLL/ goes through an intermediate [RLL] stage en route to [RRL], are impossible. A simultaneous mode of operation is also tricky. If the rule is caused by avoidance of LL sequences, why shouldn't a /LLL/ string just become [L R L], instead of the less faithful [R R L]?

So are the facts in (4-5) really true? In (4), the L-to-R sweeps are reported at slow speeds, when no flattening out takes place. If this is correct, then there is no way out: the grammar must be able to explain these sweeps, even in the absence of left-branching cyclic structure. (5) involves fast speech, where the sequences of rises are said to flatten out. In order to investigate the question of whether there is any real evidence for the existence of these L-to-R sweeps, we conducted two phonetic studies. Specifically, we wanted to answer two questions:

1. Do speakers really produce L-to-R sweeps even in unstructured (non-cyclic)strings?
2. As speech rate increases, are the rising targets still detectable, or are the originally L tones now phonologically H, with H targets?

2. Pilot study

Kuo and Yip 2004 report on a pilot study on strings of three to five underlying low tones of two types: left-branching strings, and unstructured strings. The results suggest two things. First, the preference for binarity is very strong, so that four-syllabled strings were normally broken into two groups of two, and five-syllabled strings into two plus three. The longest grouping was thus one of three lows together if the entire utterance was an odd-parity string. Second, very few of these showed clear evidence of surface L-to-R sweeps, and only at slow speeds. At higher speeds, they were so flattened that they looked (and sounded) like HH strings. However, the stimuli were not designed so as to allow for the kind of careful analysis that would demonstrate conclusively whether there were in fact rising targets still present in these apparently high spans, so we undertook a second and more carefully designed study.

3. Experiment

The model we assume has a phonological component that results in a representation with output tones on each syllable. These become the pitch targets in the phonetic component. However, they may be substantially deformed by the limitations of the articulatory system, particularly at higher speech rates, so that identifying the targets becomes quite delicate.

We conducted a phonetic experiment on long strings of L tones, with the goal of establishing the pitch targets for each syllable at different speech rates. To test for the reality of L to R sweeps, we used stimuli consisting of strings of two to four digits followed by the syllable [youL]. Including this syllable, the longest strings have five L syllables in a row, and could thus yield a maximum of four rises, RRRRL. The speakers were asked to say these at three speech rates, fast, medium, and slow. In the pitch tracks we do not show the final [youL], since it does not change, and we will parenthesize it in the outputs in the text.

3.1Stimuli and recording procedure

3.1.1 Stimuli

The stimuli shown in (9) consist of nine question and answer pairs. The questions take the form X youL meiR youL? '*Is there X?'* where X consists of 2-4 identical number names. Three number names are used, yaoH '*one*', lingR '*zero*' and liangL '*two*'. YaoH and liangL are alternatives of the more commonly used yiH and erF, respectively, chosen so as to provide visual landmarks for labelling the data, as explained in section 3.2 below.

(9)

H series:	
LingR yaoH yaoH yaoH youL meiR youL?	LingR yaoH yaoH yaoH youL.
LingR yaoH yaoH youL meiR youL?	LingR yaoH yaoH youL.
LingR yaoH youL meiR youL?	LingR yaoH youL.
R series:	
LingR lingR lingR lingR youL meiR youL?	LingR lingR lingR lingR youL.
LingR lingR lingR youL meiR youL?	LingR lingR lingR youL.
LingR lingR youL meiR youL?	LingR lingR youL.
L series:	
LiangL liangL liangL liangL youL meiR youL?	LiangL liangL liangL liangL youL.
LiangL liangL liangL youL meiR youL?	LiangL liangL liangL youL.
LiangL liangL youL meiR youL?	LiangL liangL youL.

The H and R series were designed to allow comparison with the possible outputs of the two tonal changes that are known to affect the L series, the sandhi rule and the high leveling effect. The recording materials also included a falling (F) series with the numeral [liuF] *'six'* and a low (L) series with [wuL] *'six'*. Their acoustic variation patterns were found to be similar to those reported in the present paper. We will not discuss them here due to space limitations.

3.1.2 Recording

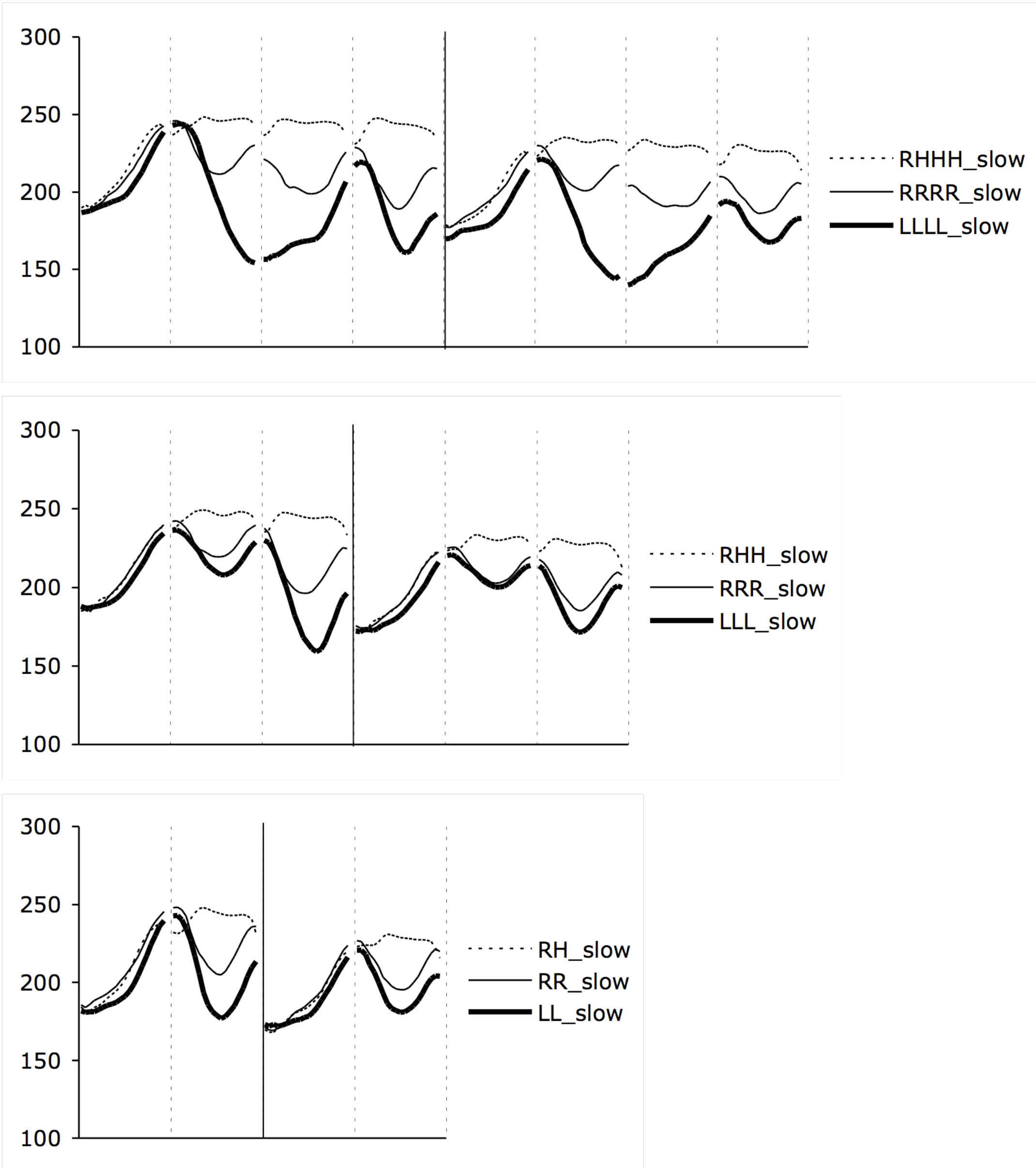
Recording was conducted in the Speech Sciences Laboratory at the Department of Phonetics and Linguistics, University College London. The subjects were seven native speakers of Standard Chinese, 4 females and 3 males. They were all born and raised in Beijing and their ages ranged from 22 to 37 years old. The question-answer pairs were printed in random order on a sheet of paper, using Arabic numerals for the digits, and a different randomization list was used for each subject. All stimuli were recorded at three speech rates (normal, fast, and slow), with three repetitions for each rate. Subjects were asked to produce the whole set of stimuli at one speech rate, and then at another rate. The order of speech rate was also randomized, and a different order was used for each subject. Before recording started, subjects were given some practice to get familiarised with the stimuli, and with speaking them at different rates. During recording, subjects were asked to repeat again whenever a sentence was not produced properly as judged by the first author who is a Standard Chinese speaker. The speech signals were recorded into a computer using the SFSWin programme (Speech Filing System, http://www.phon.ucl.ac.uk/resource/sfs) at a sampling rate of 44.1 kHz.

3.2 Data Analysis: F0 Extraction for visual inspection

The extraction of F0 contours was done with a procedure that combines the vocal cycle marking of the Praat program (www.praat.org), manual rectification and time-normalization. First the waveform, spectrogram and vocal cycle markings of each sentence are displayed on a computer screen using a Praat script. Then the vocal-cycle markings were manually corrected for errors such as missed vocal cycles and double markings. Following that, boundary labels were manually inserted at locations of visually recognizable landmarks. For /l/, the boundary label was placed at the onset of the /l/ murmur, where the intensity of all formants abruptly lowers due to the raising of the tongue tip against the alveolar ridge. For /y/ the boundary label was placed at the point where F2 reaches a peak and starts to drop. The Praat script then converted the markings into F0 values and removed all local spikes in the F0 curves using a trimming algorithm developed in our previous research (Xu 1999). A custom-written Perl program then converted the trimmed F0 into time-normalized contours by getting the same number of evenly distributed F0 points for each syllable. The Perl program then averages over the three repetitions produced by each speaker to generate a mean time-normalized F0 contour. Finally, these contours were read into Microsoft Excel to make the plots presented in the present paper. For the plots in which the F0 contours are averaged across several speakers, F0 was first converted to a logarithmic scale before averaging and then back to Hz after averaging.

3.3 Discussion

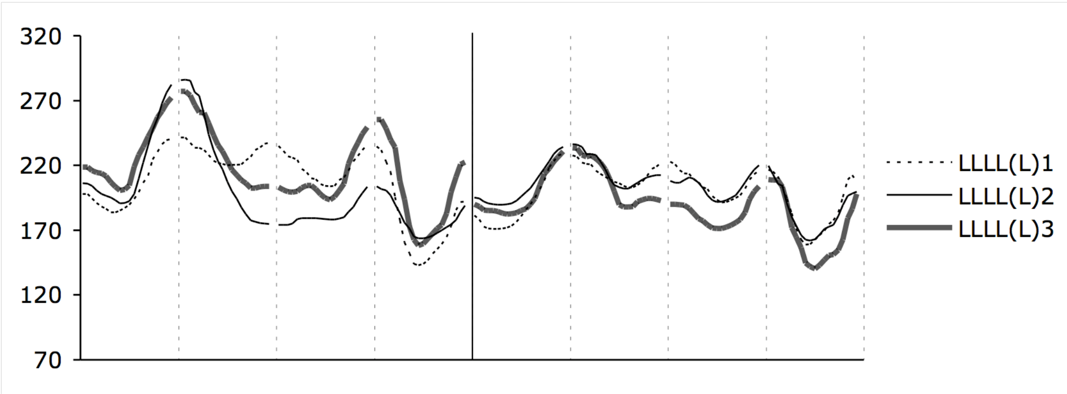
Starting with the slow speech, in slow speech all subjects show signs of L to R sweeps for the two and three digit strings, so that /LL(L)/ > [RR(L)] and /LLL (L)/ > [RRR(L)]. These can be seen in the second and third plots in Figure 1. These same plots also show clearly that the R targets for the underlying L tones are similar in shape to those for an underlying R, but slightly lower in pitch. For the four-digit strings, shown in the first plot in Fig. 1, most subjects break the digits up into two spans: [RL][RR(L)], showing the strong effect of binarity.



Normalized time

Figure 1. Overlaid mean F0 *tracks of H, R and L sequences said at slow rate by three speakers who produced all LLLL(L) sequences as RLRR(L). Each curve is a time-normalized average of 9 trials, 3 by each speaker. In each panel, the left half shows the* F0 *tracks of the questions; the right half shows* F0 *tracks of the answers. The dashed vertical lines mark the syllable boundaries.*

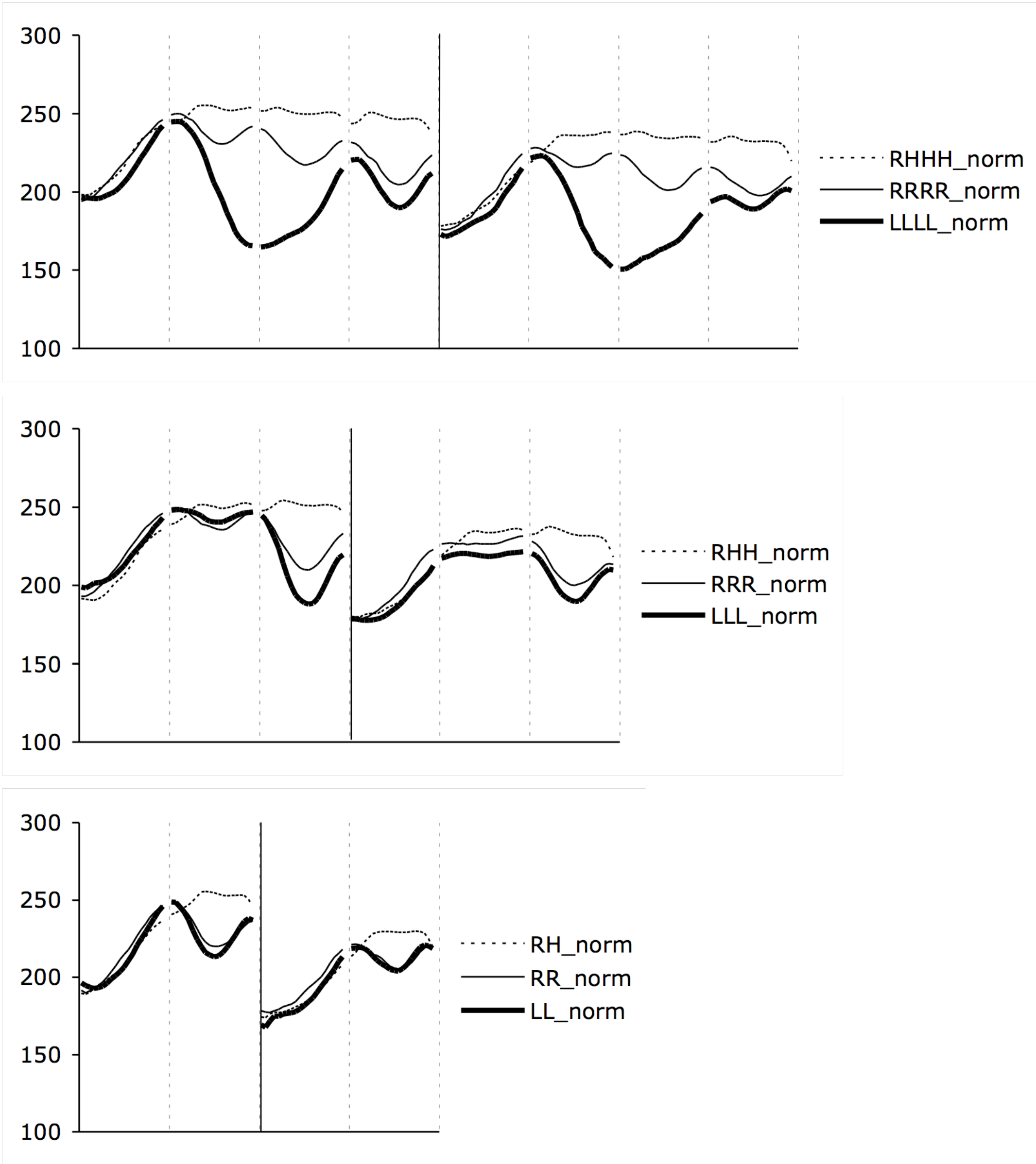
Although most subjects broke up the four-syllabled strings into two binary domains, one subject had some cases in which she produced a [RRRR(L)] sweep (4 out of 6 trials). The figure below shows three separate trials from this speaker. The facts are particularly easy to see by inspection of the pitch tracks for the first four syllables, which show the question portion of the data. In LLLL(L)1 (the dotted line), every syllable is clearly rising, whereas in the other two utterances the second syllable is L, as it was for the other speakers.



Normalized time

Figure 2. Individual F0 tracks of subject 4's LLLL(L) production at slow rate. The left half shows the F0 tracks of the questions. The right graph shows those of answers. The dashed vertical lines mark the syllable boundaries.

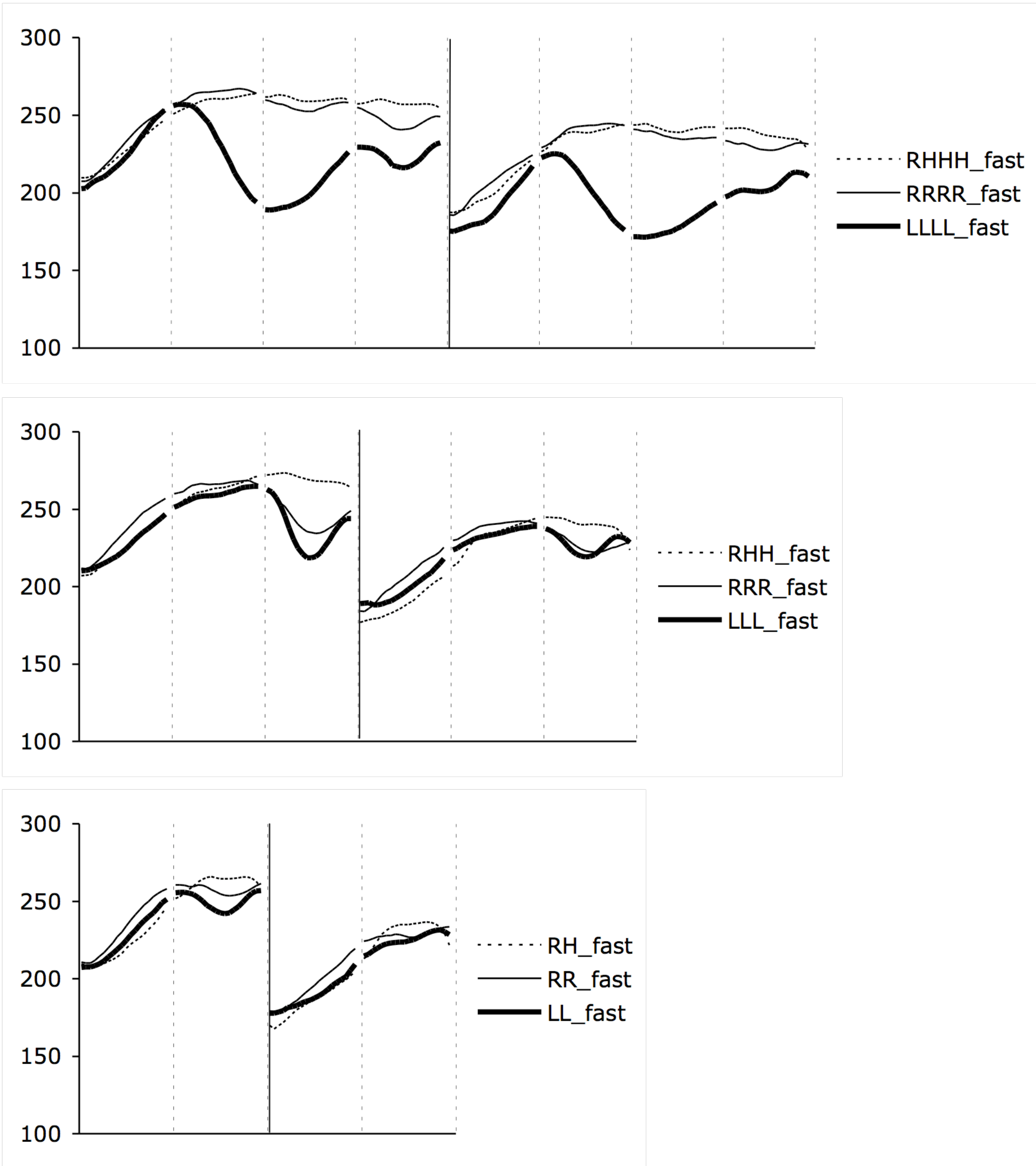
At normal speeds, shown in Figure 3, the two digit strings show the same L to R sweeps as at slow speed. For the three digit strings, the same is true, but we begin to get signs of a flattening effect on the second syllable as its duration becomes too short for a full pitch movement to be achieved. Crucially, the amount of flattening appears to be identical to the flattening observed with underlying /RRR(L)/ strings, suggesting that the effect is not one of any further phonological change in targets, but a phonetic effect due to the inertia of the articulators (Xu & Sun, 2002). Supporting this view, the flattened RR sequences remain distinct from a HH sequence, and still show small vestiges of their R targets. For the four-digit strings, most subjects again break the span into two, giving [RL][RR(L)]. As a result the third syllable, which has a rising target, looks very different from the third syllable of an underlying rise, because it has to rise all the way from the preceding L. The overall picture, then, is that the pitch targets at both slow and normal speeds are identical, but at normal speed some flattening is beginning to take place.



Normalized time

Figure 3. Overlaid mean F0 *tracks of H, R and L sequences said at normal rate by three speakers who produced all LLLL(L) sequences as RLRR(L). Each curve is a time-normalized average of 9 trials, 3 by each speaker. In each panel, the left half shows the* F0 *tracks of the questions; the right half shows* F0 *tracks of the answers. The dashed vertical lines mark the syllable boundaries.*

At fast speeds, the two and three digit strings show an increased degree of flattening, with the underlying LLL strings and RRR strings behaving almost identically to each other except for the slight overall pitch difference. This can be seen in the lower two plots in Figure 4. The most interesting are the four-digit strings, shown in the first plot of Fig. 4. For speakers who break them into two domains at slow speeds, they continue to do this at fast speeds, so the flattening out that is so striking for the underlying RRRR strings is only visible in the final two syllables of the underlying LLLL strings, now with RLRR targets.



Normalized time

Figure 4. Overlaid mean F0 *tracks of H, R and L sequences said at fast rate by three speakers who produced all LLLL(L) sequences as RLRR(L). Each curve is a time-normalized average of 9 trials, 3 by each speaker. In each panel, the left half shows the* F0 *tracks of the questions; the right half shows* F0 *tracks of the answers. The dashed vertical lines mark the syllable boundaries.*

Not surprisingly, the speaker who at slow speech rates shows RRRR(L) targets, at fast speech rates shows pitch contours much more similar to the underlying RRRR(L) cases, with an almost completely flat pitch contour on the medial syllables, as can be seen in the final plot below in Fig 5. Each portion of this figure shows her four syllabled strings at one of the three pitch rates:

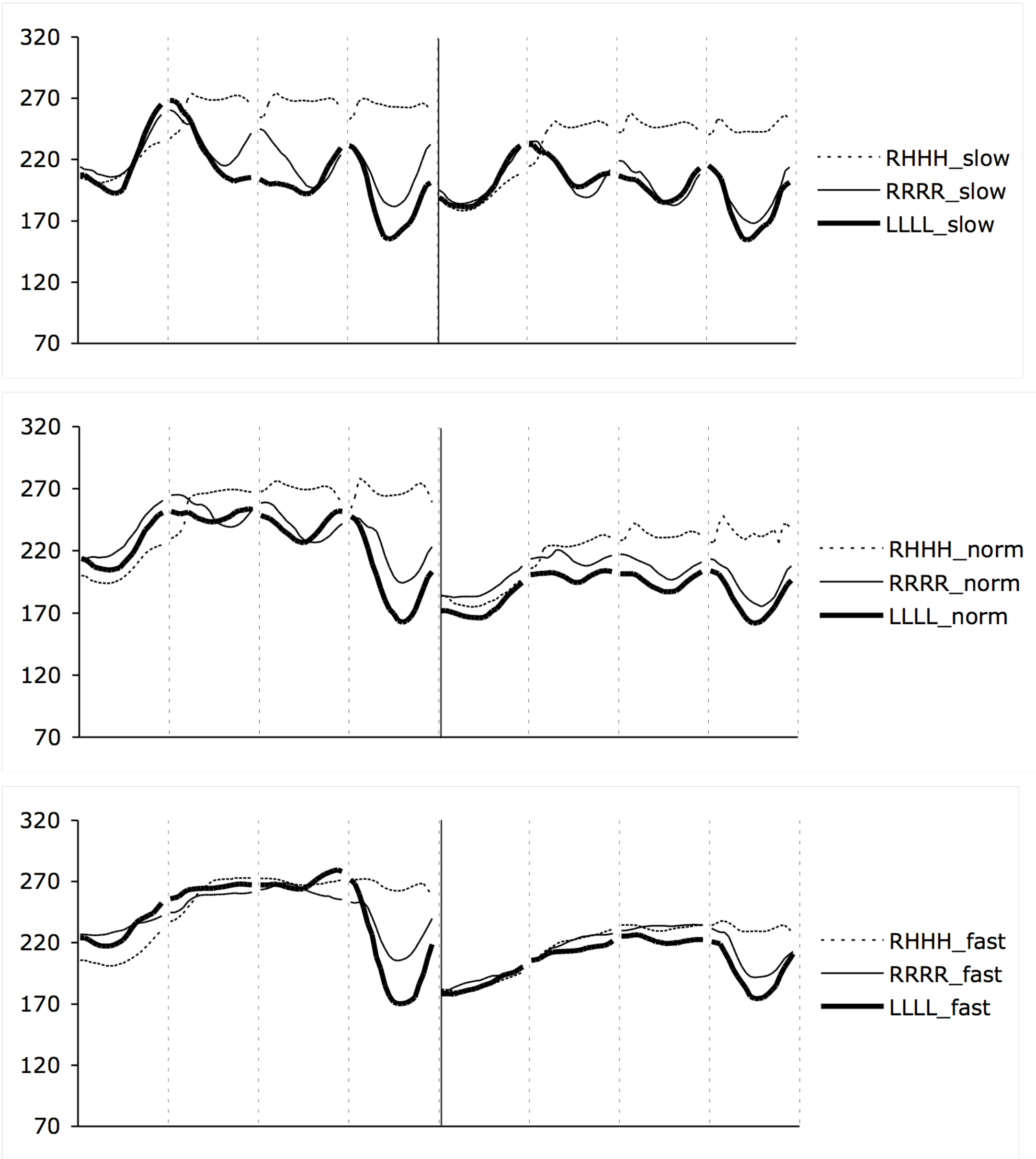


Figure 5. Overlaid mean F0 tracks of H, R and L sequences said at fast rate by subject 4 who produced most LLLL(L) sequences as RRRR(L). Each curve is a time-normalized average of 3 trials.

The data also strongly suggest that the 'sandhi rule' that is reported to change RRRR strings to RHHR is not a phonological rule that changes the pitch targets, but a matter of phonetic realization. As speech speeds up, the rising sequences gradually flatten out, and there is no discontinuity between the RRRR stage and a RHHR stage. A second fact supports this view: the resulting flat contours are not necessarily identical, with the ones that are underlyingly RRRR remaining lower in pitch than the ones that are underlyingly RHHH.

We looked in more detail at the flattening effect and its relationship to speech rate. A regression analysis with syllable duration as predictor and depth of F0 valley within a syllable as dependent variable showed that the two are positively related, with $R^2 = 0.26$ for both R and L, as can be seen in Figure 6. The valley depth was measured as the difference between F0 minimum to the maximum in the syllable, and the latter was either before or after the former.

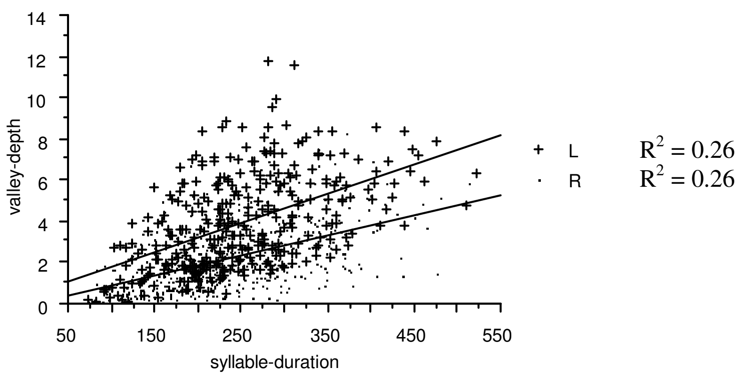


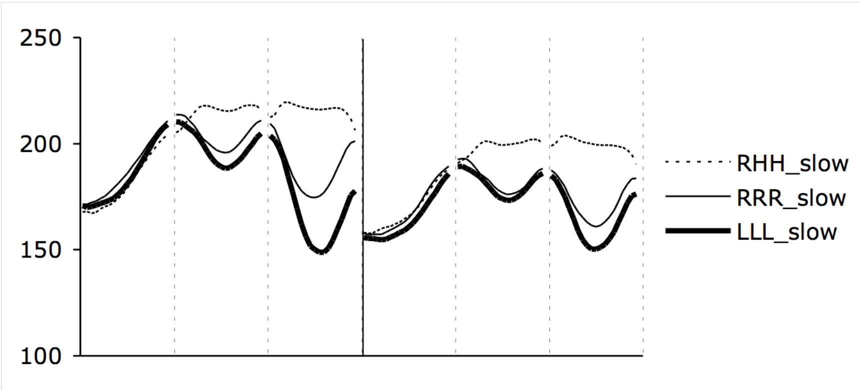
Figure 6. Variation of F0 valley depth (measured in semitone) as functions of duration of the second, third and fourth syllables in the RR(L), RRR(L), RRRR(L), LL(L), LLL(L), and LLLL(L) series. Syllables where F0 contours were fully flattened or had reverse contours (rise-fall) were excluded. As a result, the plot includes 357 observations from slow rate, 281 from normal rate and 189 from fast rate.

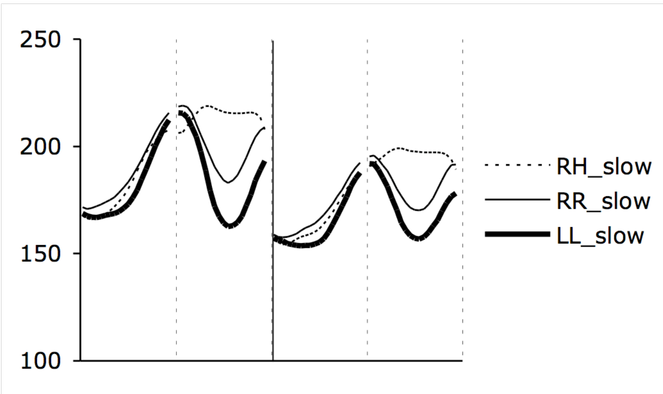
Further analyses revealed that, as shown in Table 1, the speed of F0 change at the normal and fast rates approached or exceeded the maximum speed of pitch change reported in Xu & Sun (2002). This means that the speakers have approached their speed limit in changing pitch and the only way to go faster is to totally flatten F0. This is further supported by the dramatic decrease in the number of F0 contours containing detectable valleys from slow to normal to fast rate: from 357 to 281 to 189 (see caption of Figure 6).

Table 1. Speed of pitch change (semitones/second) at three speech rates as compared to the maximum speed of pitch change reported in Xu & Sun (2002). The rise speed was measured from the F0 drop from the syllable onset to the F0 valley around the centre of the syllable and the rise speed measured from the F0 valley to the syllable offset. Those from Xu & Sun were taken from Table VI of the paper.

		Rise (4 st)	**Fall (4 st)**
Max rate (Xu & Sun 2002)		33.2	33.7
	Slow rate	22.5	24.8
Present study	Normal rate	30.7	29.2
	Fast rate	40.1	35.6

Although it was not designed for this purpose, one of the robust findings of the acoustic work in this study is that underlying R tones and the ones that result from tone sandhi in fact remain distinct: they have the same shape, but the sandhi tone is slightly lower in pitch throughout. If one looks at the pitch tracks in Figure 7, which plots the averaged F0 contours by all seven subjects, and if we look particularly at the second syllable of the RR and RL cases, which are now supposedly RR and RR, the second syllable that was an underlying L clearly has a much lower initial target than the underlying R. These data agree with similar findings by Peng (2000) and Xu (1997).





Normalized time

Figure 7. Overlaid mean F0 tracks of H, R and L sequences said at slow rate by all seven speakers. Each curve is a time-normalized average of 21 trials, 3 by each speaker. The averaging of F0 across all speakers in these cases is justified by the fact that LL(L) and LLL(L) were produced as RR(L) and RRR(L) by all of them.

In the model we are using here, the acoustic difference implies a difference in the phonological targets. Suppose then that the rising tones produced by sandhi are LH, whereas the underlying rising tones are in fact MH tones, not LH, (and underlying falling are HM, not HL). They are then phonologically distinct, a hypothesis that will play a key role in the formal analysis below.

4. The phonological analysis

The results of this study show clearly that the phonology must be able to do three things.

First, it must allow for cyclicity or some comparable mechanism by which at normal speeds structured left-branching strings can convert [[[LL]L]L] into [[[RR]R]L], but right-branching strings [L[L[LL]]] can only become [R[L[RL]]]. The literature is very clear on this point, as the example in (1) shows, so in this study we only looked at unstructured and left-branching strings, and we will not offer an OT analysis of the cyclic cases, instead referring the reader to the extensive literature on cyclicity in OT, including Duanmu (1997), McCarthy (2002:170, 184-5) and references therein.

Second, the analysis must factor in binarity, so that an alternative output for unstructured [LLLL] is [RL][RL], but [LLL] stays one domain [RRL]. The overwhelming cross-linguistic evidence for binarity as a phonological mechanism makes it unlikely that this effect is phonetic, and suggests a role for prosody in creating binary foot-like domains.

Third, it must allow for some mechanism by which unstructured strings can convert [LLLL] into [RRRL]. [ii]This is the main focus of the OT analysis.

The results also show clearly that phonology does not turn the R targets (whether underlying or from a LL sequence) into H targets: this is done by the phonetics.

While there is evidence that Standard Chinese rises are dynamic phonetic targets in themselves rather than interpolations between high and low targets (Xu 1998, Xu and Wang 2001), the phonological analysis below follows the usual practice in phonology, and presupposes that they are represented as sequences of tones, LH, rather than as unitary elements, R. A later interpretive process may of course convert the LH to an R pitch target. Henceforth, then, all tones will be reduced to L, M and H primitives, so that R= MH or LH (depending on whether it is underlying or derived), and F=HM.

Before moving to the formal part of the analysis, we need to look briefly at the interaction between the sandhi rule and prosodic factors, in order to understand the motivation for the sandhi rule itself, and the binarity effects.

4.1 Background on tone-prominence interaction

We will argue that one of the major factors responsible for the form of the sandhi rule is an avoidance of solitary low tones on prominent syllables. In this section we discuss the evidence for prominence in Chinese, whether the target of sandhi is a prominent position or not, and the effects of contrastive stress.

The question of whether or not Standard Chinese has stress is controversial (see Chen 2000:286ff for a useful summary). There is often no clear difference in the stress level of adjacent syllables, including the majority of cases where L tones change to R before another L. For example /laoL xiaoL/ *'old and young'* becomes [laoR xiaoL]. There are nonetheless some situations where stress differences are clear, and it can then be seen that main stress may be on any fully-toned syllable, including both L and R, as shown by the following examples, where the second syllable, a perfective marker 'le', is toneless and reduced, and therefore clearly unstressed (Here and throughout stress is shown by underscoring.): maiL-le *'bought'*, laiR-le *'came'*.

What about contrastive stress? Contrastive stress is realized not only by changes in the stressed item itself, but also by the suppressed pitch range of all the subsequent syllables in the utterance. If necessary, even a L-tone syllable can bear contrastive stress. As reported in

Xu (1999a), when native Mandarin speakers put focus on a non-sandhi L they do it by making its F0 even lower while also lowering the pitch range of the subsequent words. However, there is evidence that speakers prefer not to place stress on L toned syllables if there is an alternative. Emphasis placement shows avoidance of a L-toned syllable (Hoa 1983:98). When an adjective phrase is contrastively stressed, the stress normally shows up on the adjective portion, as in henL zhongF *'very heavy'*, taiF zhongF '*too heavy*'. However, when the adjective is L toned, stress shifts to the modifier, as in taiF xiaoL *'too small'*. If this modifier is itself underlyingly L, since it is in front of another L it will undergo the sandhi rule, and become rising R: henR xiaoL *'very small'*. A summary of the possibilities is shown below, where X stands for any tone except L:

/X X/ → X X
/L X/ → L X
/X L/ → X L
/L L/ → R L

As a result the outputs contain no contrastively stressed L-toned syllables.

A second piece of evidence pointing in the same direction comes from a change of phrasing under contrastive focus. When two L toned syllables are separated by a phrase boundary, sandhi cannot apply, as in the verb [maiL] *'buy'*, in (6a) below. However, if the first of these syllables is placed in a contrastive focus context, the phrasing is changed to group the two L syllables together, as in (6b). This has the result that the focussed element [mai] *'buy'* is now R, not L: (Zhang 1988, Shih 1997:112)

(6)		*a. Normal phrasing*:		b. *Phrasing under contrastive focus*:	
		only buy stocks		only [buy stocks	not sell stocks
		zhi mai gu-piao		*zhi [mai gu-piao,*	*bu mai gu-piao*
	U.R.	L L L F	U.R.	L [L L F	R F L F
	Sandhi	(R L)(L F)	Sandhi	(L) [(R L) (F)	

If the following syllable is not L, the sandhi rule cannot of course apply, in which case the focused element undergoes the changes discussed in Xu (1999a) and earlier in this section, and some reports suggest that it may even become the 'full third tone', which is still low, but longer and with a rise, and usually only found pre-pausally. See also Shen (1990: 51).

These observations suggest a possible reason for the sandhi rule to exist at all. Suppose it is the head of the prosodic unit (preferentially but not necessarily binary, see Shih 1986, Chen 2000) within which sandhi takes place. It is known that cross-linguistically heads both avoid L and prefer to be H (De Lacy 2003), and from this perspective the change from L to R improves the tonal quality of the head syllable. However, the evidence bearing on the position of the head in Standard Chinese is mixed, given that there is no overt stress in the language. There are four pieces of evidence in favour of the prosodic unit being left-headed. First, true unstressed syllables are only found non-initially, so that in a sequence of one toned syllable followed by one toneless syllable, the structure is unequivocally head-first. See Duanmu (2000:252) for discussion. Second, Chen (2003) shows that the first and third syllables of a four-syllabled word lengthen more than the second syllable under word focus. The word-final syllable lengthens most, showing a word edge effect, but Chen argues that the unequal patterns of lengthening on the non-final syllables support a trochaic analysis in which the structure is (SW) (SW). Third, Kochanski et al (2003) show that phonetic strength measurements support a S-W metrical pattern. Fourth, and returning to sandhi, it is clear that sandhi can take place on a head syllable in some circumstances. In the following reduplicated phrases the input has L tones on both syllables, and the output has a sandhi R tone on the first syllable, but no tone on the second, unstressed syllable (examples from Wee 2004:49)[iii]:

(7)	*zouLzouL*	→	*zouR zou*	'take a short walk'
	xiaoLxiaoL	→	*xiaoR xiao*	'smallish'

A somewhat different example comes from code-switching. When an English word is introduced into a Chinese sentence, English unstressed syllables with low pitch trigger sandhi. In (8a) below, the first syllable of '*conversation*' is stressed, with a falling pitch, and the L tone on the preceding Chinese word is unchanged. However, in (8b) the first syllable of '*connection*' is unstressed, with low pitch, and it triggers the L-to-R sandhi on the preceding Chinese syllable.

(8)	a. *haoL cònversátion*	vs. b. *haoR connéction*
	good conversation	good connection

This is the normal Chinese tone change found in a sequence of two L tones. The interest of the data in the present context has to do with the stress facts. Since the triggering syllable [kə] of '*connection*' is clearly unstressed, the LL sequence cannot possibly have a WS stress pattern but rather the sequence must be SW, and the target syllable must presumably be stressed (Chang 1992:197).

Moving to larger prosodic domains, there is some phonetic evidence that phrase-initial syllables are prominent. Li (2003:85) reports the following indicators. Yan and Lin (1988, cited in Duanmu 2000) find that they have the greatest pitch range. Yang (1992) finds they have the longest onset duration, greatest amplitude, and highest F0 peaks. As pointed out by a reviewer, this could however be simply domain-initial strengthening (Fougeron and Keating 1997, Keating et al 2004), not evidence of stress.

Whatever the headedness of utterance internal constituents, there is little doubt that for the utterance as a whole, the final syllable is the head. These syllables show several special properties, in particular they are longer, and a L in this position may show up with a final low-rise ('full third tone'). Some authors also claim they have a greater pitch range, and they show the highest H of their phrase, See Shen (1990:60), Hoa (1983), Wee (2001:171) for discussion.

We are left with a rather complex picture. At the utterance or perhaps major phrase level, final syllables are prominent. Within these, however, there is some evidence for left-headed structures, suggesting that the targeted syllable in a LL sandhi domain is the head (whether or not it has any overt manifestations of stress). If this is right, it means that the pressure to avoid L-toned heads, together with a prohibition on LL sequences, outweighs the preference for leaving heads unchanged. We will return to this in the analysis.

4.2 An OT proposal for sequences of two tones

We start with a mini-grammar that explains why LL sequences are resolved by changing the first L to a rise, rather than to any other possible output. It makes use of three main elements. First, like many others, we assume that LL violates the Obligatory Contour Principle (or OCP; Leben 1973, Myers 1997), and that Standard Chinese observes OCP-L within a prosodic constituent. Given the assumption about the representation of the derived rise as LH (and a possible derived fall as HL), the OCP will be violated by L.LH or HL.L as much as it is by L.L. Second, it relies on the finding that the first L is the head, and that L on heads is avoided. This insight is captured using the positional markedness constraint *HEAD/L, a constraint taken from De Lacy (2003), which expresses the rejection of L tones on prominent syllables.[iv] This constraint must be interpreted here as meaning that head syllables may not have a simple L tone, and is not violated by a syllable with a LH sequence. An alternative, suggested by the editors, would be to use a positive constraint stating that heads must have a H (see also Yip 2002:205). [v] Third, it assumes that input tones are not deleted. If the rise is indeed LH, then the change from input L to sandhi LH does not delete the L, but adds a H.

We start by looking at the simplest case of a sequence of two low tones. In tableau (9) below, we compare the unchanged input (9d) with three ways of obeying the OCP. The OCP rules out candidate (9d), with the unchanged input LL sequence. Candidate (9c) is ruled out by MAX-T, because it has deleted one input L. Candidates (9a-b) satisfy both these constraints, and they tie on the constraint against inserting tones, DEP-T, but candidate (9b) has a L tone on the head syllable, in violation of *HEAD/L, so candidate (9a) wins. This tableau establishes the ranking OCP-L>> DEP-T. Other rankings will be established shortly.

(9) *Tone sandhi in a /LL/ string*

/L.L/	OCP-L	MAX-T	DEP-T	*HEAD/L
☞ a. (<u>LH</u>.L)			*	
b. (<u>L</u>.HL)			*	*!
c. (<u>H</u>.L)		*!	*	
d. (<u>L</u>.L)	*!			*

Notice that one other possible solution HL.L would still violate the OCP-L because of two L tones in a row, even though one of them is a component of a contour tone. We will see further evidence later that this is correct. Finally, note that the only lexical L tone is the low level tone, since the lexical rise and fall are MH are HM respectively, so the only source of underlying OCP-L violations comes from sequences of low level tones.

It is important to observe that although *HEAD/L plays a major role in the grammar, only the OCP-L can force tonal change to take place. The grammar below makes this point clear: if MAX-T and DEP-T are ranked above *HEAD/L, L tones will survive on heads in the absence of an OCP-L violation. The tableau below illustrates the case of L before an underlying MH.

(10) *Survival of L on heads when followed by non-low tones*

/L.MH/	OCP-L	MAX-T	DEP-T	*HEAD/L
☞ a. (L.MH)				*
b. (LH.MH)			*!	
c. (H.MH)		*!	*	

In conjunction with the earlier results, we now have the rankings OCP-T >> DEP-T, and MAX-T, DEP-T >> *HEAD/L.

We now turn to the interaction of sandhi with contrastive focus. Recall that focus can force re-phrasing so that the focused word starts a constituent (Shih 1997: 112), something also observed in Greek (Condoravdi 1990). Since constituents are left-headed, the focused word is now a head, and undergoes sandhi so that it is LH. We view the re-phrasing as triggered by the desire to avoid L on the focused element, and so we will introduce a slight variant of *HEAD/L, which we will call *FOCUS/L. This must dominate BINARITY, the constraint which normally prefers binary constituents, and it probably also dominates some of the other constraints that align morpho-syntactic and prosodic structure, but these issues are beyond the scope of this paper. In the tableau below we use double underlining to mark the focused element. The parentheses show the prosodic phrasing, not the syntactic phrasing, following the same practice as in (6) above.

(11) Change in phrasing under focus avoids L on focused element

/L.L.L.HM/	*FOCUS/L	BIN
☞ a. (L)(LH.L)(HM)		**
b. (LH.L)(L.ML)	*!	

Provided *FOCUS/L >> BIN, the normal preference for binary constituents will be over-ridden in the focus condition, and candidate (a) will beat candidate (b). In other languages, focused elements may be moved to positions where they can receive prominence, or H tones may be inserted on focused elements. Here we see a subtler phenomenon: an independently motivated phonological process of tone sandhi is being put to use to improve the tonal properties of the focused element. This tone sandhi process is only triggered by the OCP banning LL sequences within the prosodic constituent, but when possible *FOCUS/L selects between two different domains of application, and thus ensures that the focused element is the target.

4.3 L-to-R sweeps: Sequences of three or more tones

We are now ready to look at the core problem with which we started: how do we account for L-to-R sweeps. We have established that such sweeps exist, and that they can create sequences of up to three rises on an unstructured string. Let us start with a LLL input, which all speakers change to RRL. Phonologically, then, we need to produce a (LH LH L) output. At first glance this is surprising, since the OCP might appear to be more economically satisfied by a single H insertion, as in L.LH.L, and there is thus no reason why LH.LH.L with *two* H insertions, should be preferred. However, recall our earlier remarks about the OCP-L applying even to component tones. The reason the LH LH L output is preferred to the L LH L output is that a L LH sequence is still an OCP-L violation at the level of the component tones. [vi] The constraint will be violated by HL.L, or L.LH, and of course L.L. In the tableau below the OCP violations are calculated at both tonal levels, and candidate (a) will correctly beat candidate (c).

(12) *A successful grammar for L-to-R sweeps*

/ L L L/	OCP-L	MAX-T	DEP-T	*HEAD/L
☞ a. LH LH L			**	
b. L H L		*!	*	*
c. L LH L	*!		*	*
d. L L L	**!			*

In this analysis, the L-to-R sweep is achieved not by left-to-right iterative rule application, but as the optimal way to avoid OCP violations at the level of either full or component tones, and achieve good tone-to-prominence mapping, with minimal faithfulness violations.

These three-syllabled domains are commonplace, arising in all odd-parity unstructured strings, and they behave entirely consistently in this way. We now turn to the longer cases, /LLLL/ inputs, where speakers vary. For most speakers, BINARITY forces these to be broken up into two domains, each of which can be handled as shown above. However, a few speakers have at least some outputs with a single domain, even at slow speeds, raising a challenge for the grammar. At fast speeds, this is reported to be quite common. As before, the penultimate syllable will have to change to LH, in deference to OCP-L and MAX-T. Below we show only the last two syllables, to make the point.

(13) *Four-syllabled strings: the last two syllables:*

/L L L L/	OCP-L	MAX-T	DEP-T	*HEAD/L
a. ....L L	*!			
b. ....H L		*!	*	
☞ c. ....LH L			*	

Once we look at the full string, and consider all candidates that satisfy both MAX and the OCP-L over these first two syllables, we can see that both candidates (14b-c) also create new illicit sequences [L.LH] that contain adjacent low component tones, and are thus ruled out. Candidate (14a), with a L-to-R sweep, is the winner.

(14) *Four-syllabled strings: the full grammar*

/L L L L/	OCP-L	MAX-T	DEP-T	*HEAD/L
☞ a. LH LH LH L			***	
b. LH L LH L	*!		**	

c. L LH LH L	*!		**	*
d. L L LH L	**!		*	*

The solution generalizes to any number of L's in sequence, turning them all to LH.

5. Conclusions

The acoustic data in this paper have shed new light on the division of labour between phonology and phonetics in tone sandhi in Standard Chinese. We have shown that the change from L to R is indeed phonological, but the flattening out of R to H is a purely phonetic effect, not a phonological rule. We have also shown that there are indeed instances in which speakers change a long sequence of L tones to a sequence of rises, in non-cyclic environments, and that the phonology must therefore be able to model this change. In a one-step output-based theory like OT, iteration cannot be handled, but we have also shown that once we recognize that the rises created by the tone sandhi process are LH , whereas the underlying rises are MH, a one-step analysis of the L-to-R sweeps is possible within OT. The analysis makes crucial use of two things. First, there is an avoidance of low-toned heads. Second, the OCP is obeyed on two levels: the level of the complete tonal complex, and the level of the constituent tones.

Acknowledgments

The three authors bring different areas of expertise to this paper. The phonetics work was conducted by Yi Xu and Yu-ching Kuo, and the phonological analysis by Moira Yip. Portions arose out of the earlier pilot study by Kuo and Yip presented at the TIE conference in Santorini, but all the acoustic data in this paper is new. We are grateful to an anonymous reviewer, the tireless editors Carlos Gussenhoven and Tomas Riad, and also to many of the participants in the TIE conferences in Santorini and in Konstanz for insightful comments, particularly Robert Ladd and Gösta Bruce. Thanks also to Lisa Cheng for advice on syntax. All errors are of course our own.

References

Beckman, Jill. 1997. Positional Faithfulness. Doctoral dissertation, University of Massachusetts, Amherst.

Chang, M-C. L. 1992. A prosodic account of tone, stress and tone sandhi in Chinese languages. PhD dissertation, U. of Hawaii. The Crane Publishing Co, Taiwan.

Chen, Y.Y. 2003. Durational adjustment under contrastive focus. To appear in *Journal of Phonetics.*

Chen, YY and Y. Xu 2004. Production of weak elements in speech - Evidence from F0 patterns of neutral tone in Standard Chinese. Ms, U. of Edinburgh and Haskins Labs.

Chen, Matthew 2000. *Tone Sandhi.* Cambridge University Press.

De Lacy,P., 2003. The interaction of tone and stress in Optimality Theory. *Phonology* 19, 1-32.

Duanmu, San. 1997. Recursive constraint evaluation in OT: Evidence from cyclic compounds in Shanghai. *Natural Language and Linguistic Theory* 15.3:465-508

Duanmu, S. 2000. *The phonology of Standard Chinese.* Oxford University Press.

Flemming, E. 2003. Speech perception in phonology, ed. by Pisoni, D. and R. Remez, Oxford: Blackwells.
Flemming, E. 2003. Contrast and perceptual distinctiveness, ed. by Hayes, B, R Kirchner, and D. Steriade, Cambridge: Cambridge University Press.
Fougeron, C. and Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. J. Acoust. Soc. Am. 101, 3728
Hale, M. and C. Reiss. 2002. Phonological accounts of microvariation in pronunciation. Paper given at 25th GLOW colloquium, Amsterdam.
Hoa, M. 1983 *L'Accentuation en Pekinois*. Editions Langages Croisés. Centre de Recherches Linguistiques sur L'Asie Orientale. Paris.
Keating, P., T. Cho, C. Fougeron, and C.S. Hsu. 2004. Domain-initial articulatory strengthening in four languages. *Papers in Laboratory Phonology VI*
Kirchner, R. 1998. An effort-based approach to consonant lenition. U.C.L.A. Ph.D. Dissertation.
Kochanski, G., C. Shih and H. Jing. 2003. Quantitative measurement of prosodic strength in Mandarin. *Speech Communication* 41: 625-645.
Kuo, Y.C. and M. Yip 2004. Tone and prominence: L-to-LH sweeps in Mandarin. Paper presented at the International Conference on Tone and Intonation, Santorini, Greece, September 2004.
Leben, W. (1973) *Suprasegmental Phonology.* PhD Dissertation, M.I.T.
Li, Z.Q. 2003. The phonetics and phonology of tone mapping n a constraint-based approach. MIT PhD Dissertation.
McCarthy, J. 2002: *Thematic Guide to Optimality Theory*. Cambridge University Press.
McCarthy, J. 2003. Comparative markedness. *Theoretical Linguistics* 29: 1-52. ROA 489-0102
Myers, S. (1997) OCP Effects in Optimality Theory. *Natural Language and Linguistic Theory*: 15.4, 847-892
Peng, S.-h. 2000. Lexical versus 'phonological' representations of Mandarin Sandhi tones. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, ed. by M. B. Broe and J. B. Pierrehumbert, 152-167. Cambridge, Cambridge University Press.
Shen, X-N., S. 1990. The prosody of Mandarin Chinese. *University of California Publications in Linguistics,* Vol 118. U. of California Press, Berkeley.
Shih, Chi-Lin 1986. *The prosodic domain of tone sandhi in Chinese,* UC San Diego PhD dissertation.
Shih, Chi-lin. 1987. The phonetics of the Chinese tonal system. Ms., ATT&T Bell Labs
Shih, Chi-lin 1997. Mandarin third tone sandhi and prosodic structure. In Wang, J. and N. Smith. *Studies in Chinese phonology.* Mouton de Gruyter. New York. 81-124.
Shih, C. and Sproat, R. 1992. Variations of the Mandarin rising tone. Proceedings of the IRCS Workshop on Prosody in Natural Speech, University of Pennsylvania Institute for Research in Cognitive Science. 92.193-200.
Speer, S. R., Shih, C., and Slowiaczek, M. L. 1989. Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. Language and Speech 32.337-54.
Sundberg, J. 1973. Data on maximum speed of pitch changes. Royal Institute of Technology, Stockholm. Speech Transmission Laboratory. *Quarterly Progress and Status Report. 4, 39-47.*

Wee, L-H. 2001 Faithfulness as percolation - Evidence from tone sandhi in Mandarin. In G. Horwood and S-K Kim , ed.s, *RuLing Papers 2: Working Papers of Rutgers University.* Rutgers U. Department of Linguistics, , New Brunswick.163-196.
Wee, L.H. 2004. Inter-tier correspondence theory. PhD Dissertation, Rutgers University ROA 654-0404
Xu, Y. 1994. Production and perception of co-articulated tones. JASA 95: 2240-2253
Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61-83.
Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* **55**: 179-203.
Xu, Y. 1999a. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27: 55-105.
Xu, Y. 1999b. F_0 peak delay: when where and why it occurs. *International Congress of Phonetic Sciences 1999.* San Francisco. pp 1881-1884.
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.
Xu, Y. and Wang, Q. E. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* **33**: 319-337.
Yip, M. 1989. Contour tones. *Phonology* 6.1: 149-174.
Yip, M. (1999) Feet, tonal reduction and speech rate at the word and phrase level in Chinese. In Rene Kager and Wim Zonneveld, eds, *Phrasal Phonology.* Nijmegen University Press, Nijmegen. Pp 171-194.
Zhang, N. 1997. The avoidance of third tone sandhi in Mandarin Chinese. *Journal of East Asian Linguistics* 6.4: 293-338.
Zhang Z. 1988. *Tone and tone sandhi in Chinese*. PhD dissertation, Ohio State U.
Zoll, C. 1998. Positional markedness, positional faithfulness, and licensing. Ms, MIT.

Endnotes

[i] The low tone has a variant with a low rise that occurs pre-pausally.

[ii] This is the maximum possible given our stimuli, but the alert reader may have noticed even longer sequences of rises. These have a different cause. In a LLLLyouL sequence, the structure is [[LLLL]youL]. When it changes to RRRRL, the first three syllables have changed in an unstructured environment, but the fourth syllable has changed across a syntactic boundary. From the point of view of the phonology, then, the changes are likely to have happened in two steps. First, the digit sequence undergoes the changes, then the final digit changes when the entire sentence is considered as a whole.

Underlying form	[LLLL] youL
First cycle	[RRRL]
Second cycle	[RRRR youL]

This is a standard cyclic analysis, and it is necessary to explain the asymmetries mentioned in the introduction caused by different syntactic structures. It is therefore the case that the longest string of rises created within an unstructured string in our data is three rises, not four, since the fourth rise is caused by a cyclic environment.

[iii] The output in these cases is opaque, raising an interesting challenge for OT that is beyond the scope of this paper.

[iv] Note by the way that a faithfulness account would not work: H is inserted on heads in response to positional *markedness* preferences (Zoll 1998), despite seriously violating positional *faithfulness* (Beckman 1997).

[v] As far as we can see, this would have no adverse consequences, and have the advantage of unifying underlying H, MH, HM and derived LH, since all contain the required H. However, we will retain the *HEAD/L formulation here.

[vi] This contrasts with Tianjin, where only the whole tonal complex is subject to the OCP. See Yip 1989 for details.