THE PHONETIC SPECIFICATION OF CONTOUR TONES: THE RISING TONE IN MANDARIN

Hyesun Cho[a] *& Edward Flemming*[b]

[a]Seoul National University, Korea; [b]Massachusetts Institute of Technology, USA
chohazel@gmail.com; flemming@mit.edu

ABSTRACT

This paper investigates the phonetic specification of contour tones through a case study of the rising tone in Mandarin. The patterns of variation in the realization of the rising tone as a function of speech rate indicate that the specification of this tone involves targets for the slope of the f0 rise, the magnitude of the rise, and the alignment of the onset and offset of the rise. These targets conflict so the realization of the tone is a compromise between them. This analysis is formalized as a quantitative model of tone realization formulated in terms of weighted constraints enforcing tone targets.

Keywords: contour tones, phonetic constraints

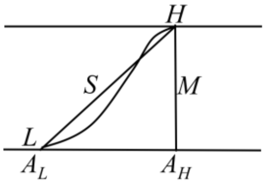
1. INTRODUCTION

In autosegmental analyses contour tones are represented as a sequence of level tones [4], and this conception has been extended to the level of phonetic implementation in many models of intonation. That is, the F_0 contour is derived through interpolation between point targets associated with individual High and Low tones, even if the tones form part of a bi-tonal accent such as a rising pitch accent [7]. On this view the transitions between tones are the result of general principles of interpolation rather than being specified by tonal targets. This is made explicit in the Segmental Anchoring Hypothesis (SAH), according to which ‘the beginning and end of a pitch movement are anchored to specific locations relative to segmental structure, while the slope and duration of the pitch movement vary according to the segmental material with which it is associated’ [6]. This approach has been successfully applied to rising pitch accents in languages such as Greek [1].

Intonational and lexical tones are represented in the same way in phonology so we might expect their phonetic realizations to be similar, but F_0 slope is an important cue to the distinction between level and rising lexical tones in a language like Mandarin, so we might expect the slope of the transition to have a specified target in such tones, contrary to the SAH [5, 10].

In this study we investigate the phonetic specification of the Mandarin rising tone, testing for targets for 3 properties of the tone (fig.1): (1) the alignment of the onset (*L*) and offset (*H*) of the rise to the segmental string, (2) the magnitude of the rise (*M*), and (3) the slope of the rise (*S*) by examining this tone under variation in speech rate.

Figure 1: A schematic illustration of a rising f0 movement.



If the onset and offset of the rise are consistently aligned to segmental anchors then as speech rate increases, moving the anchors closer together, the duration of the rise should decrease. If speakers try to maintain a target magnitude of F_0 rise then a decrease in rise duration will result in increased slope, whereas if they try to maintain a constant slope it will result in a rise of smaller magnitude. If speakers try to keep both slope and magnitude constant, then variation in the duration of the rise must be limited and consistent segmental anchoring will not be possible. Thus the patterns of variation with speech rate can reveal the nature of the targets of the rising tone.

2. EXPERIMENT

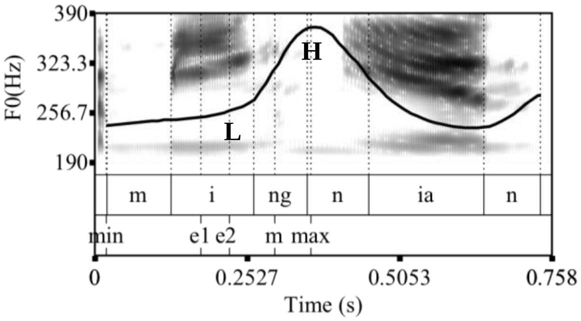
2.1. Speech materials and methods

Four speakers of Beijing Mandarin Chinese, two male and two female, were recorded reading 15 disyllabic words containing almost all sonorant consonants, where each syllable carries the rising tone. The analysis focuses on the first rising tone. The words were produced in a carrier phrase that placed the target word after a low tone. The subjects read the materials, together with filler sentences, at normal, fast and slow rates.

2.2. Measurements

The F_0 trajectory of the rising tones generally shows a relatively level interval followed by a rise (fig.2). *H* is taken to correspond to the F_0 maximum, but *L* does not correspond to an F_0 minimum, particularly where the low plateau shows a slight rise, as in fig. 2 (cf. [9]). Intuitively, *L* is at the 'elbow' between the plateau and the rise. This point was located algorithmically, adapting a procedure described in [8]: the F_0 trajectory from F_0 minimum to F_0 maximum was approximated by three straight lines, fitted to minimize squared error. The onset of the rise is the beginning of the steepest line segment ('e2' in fig.2). The timing of *L* and *H* were measured relative to word onset. The magnitude *M* is then F_0 at *H* minus F_0 at *L*, and the average slope is *M* divided by the duration between *L* and *H*. The peak velocity of the rise ('m' in fig.2) was measured by fitting cubic smoothing splines to the F_0 trajectory between the F_0 minimum and maximum, and finding the peak of the derivative of the smoothed curve. However peak velocity is highly correlated with average slope (r^2=0.92), so we will only report on the latter here. Segment boundaries in the target word were labeled manually.

Figure 2: Pitch track of the word *mingnian* 'next year'.



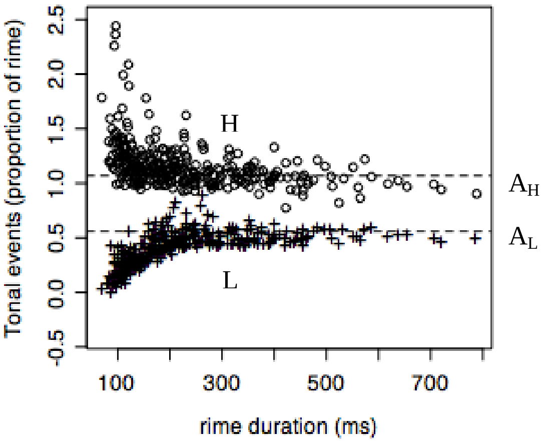
3. RESULTS

The speech rate manipulation successfully elicited a wide and continuous range of syllable durations (fig. 3). Even the fastest speech rate was judged by a native speaker to be clearly comprehensible.

3.1. Segmental anchoring

Anchor points for *L* and *H* were identified by looking for segmentally defined points that had the smallest squared deviation from the f0 events across all syllable durations. Segment boundaries and proportions of syllable and rime duration were tested. The best anchor for *L*, A_L, was about half way through the rime (56% of rime duration) and the anchor for *H*, A_H, was at 107% of rime duration, i.e. just after the end of the syllable (cf. [2, 9]).

Figure 3: Timing of H (circles) and L (+) as proportions of rime duration plotted against rime duration (ms).



The tones remain close to these anchor points, as predicted by the SAH, but there are systematic deviations from these alignment targets as a function of speech rate (fig. 3). The figure shows the timing of *L* and *H* as proportions of the syllable rime plotted against rime duration. The dashed lines mark A_L and A_H. *H* occurs progressively later than A_H (above the dashed line) as speech rate increases (i.e. as rime duration decreases), whereas *L* occurs progressively earlier than A_L as speech rate increases. That is, speakers are deviating from the anchors to avoid the rise becoming too short. This is expected if speakers have targets for slope and magnitude since maintaining relatively constant slope and magnitude implies limiting variation in rise duration.

3.2. Magnitude and slope of the rise

Neither slope nor rise magnitude is actually constant. Magnitude increases with increasing rise duration (fig.4), but not sufficiently to maintain a constant slope, so slope decreases with increasing rise duration (fig.5).

The effect of duration on magnitude is significant: a linear mixed effects model predicting rise magnitude as a function of duration, with random intercepts by speaker, fits significantly better than a model according to which rise magnitude is constant ($\chi^2(1)$=111, p<0.0001). A similar test shows that the effect of duration on slope is also significant ($\chi^2(1)$=55, p<0.0001).

To sum up: none of the properties of the rise are invariant across rates, and we observe all of the patterns of variation outlined in section 1, indicating that there are targets for all three properties of the rise. It is not possible to realize all of these targets and this conflict is resolved by a compromise between them: as A_H and A_L move closer to each other at faster speech rates, so do L and H, but the tones lag behind their anchors, so slope increases and magnitude decreases.

Figure 4: Rise magnitude plotted against rise duration. The line shows the model of this relationship discussed in section 4.

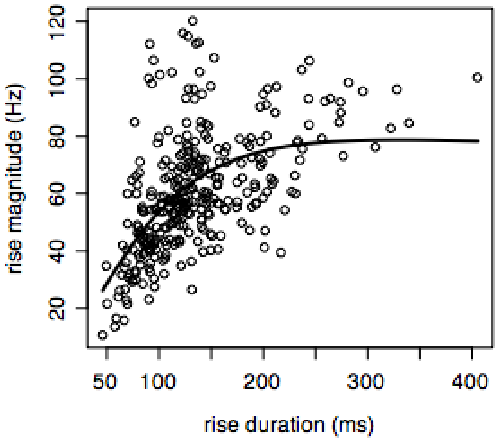
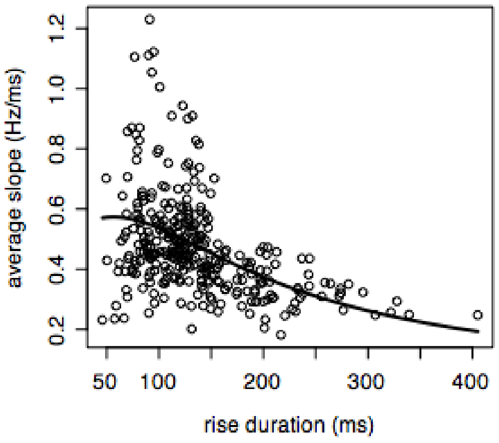


Figure 5: Average slope of the rise plotted against rise duration. The line shows the model of this relationship in (4).



4. THE MODEL

The notion that the realization of the rising tone is a compromise between violable targets for segmental anchoring of the onset and offset of the rise and for slope and magnitude of the rise can be made precise in terms of a model based on weighted constraints [3]. The four targets are enforced by the constraints listed in table 1. It is generally not possible to achieve all the targets, so the realization of the rising tone, given a particular syllable duration, is selected to minimize violation of the constraints. The cost of violation of a constraint is equal to the square of the deviation from the target (table 1), and the total cost of a candidate set of values for the timing of L and H and the average slope, S, is the weighted sum of its constraint violations (1). Minimizing this cost yields a compromise between the various targets.

Table 1: Constraints and costs of violations.

Target	Constraint	Cost of violation
Magnitude	$M = T_M$	$w_E(S(H\text{-}L)\text{-}T_M)^2$
Slope	$S = T_S$	$w_S(S\text{-}T_S)^2$
L alignment	$L = A_L$	$w_L(L\text{-}A_L)^2$
H alignment	$H = A_H$	$w_H(H\text{-}A_H)^2$

(1) $\text{Cost} = w_E(S(H\text{-}L) - T_M)^2 + w_S(S - T_S)^2 + w_L(L - A_L)^2 + w_H(H - A_H)^2$

The constraints specify targets for M, S, and timing of L and H, but we select values of just L, H and S because M can be calculated from these quantities as $S(H\text{-}L)$ (slope times rise duration). We will refer to the duration $H\text{-}L$ as D.

The minimum cost lies at the point where the partial derivatives of the cost function (1) are equal to zero. Accordingly we can derive the relationships (2)-(4) from the partial derivatives with respect to L, H and S. In (2), T_M/S is the duration that would yield a rise magnitude of T_M given a slope of S, so (2) states that the timing of L is a weighted average of its target, A_L, and the timing that would satisfy the magnitude constraint, where the weighting of A_L decreases as S^2 increases. Similarly, (3) states that the timing of H is a weighted average of A_H and the timing that would satisfy the magnitude constraint.

(2) $$L = \frac{w_M S^2\left(H - \frac{T_M}{S}\right) + w_L A_L}{w_M S^2 + w_L}$$

(3) $$H = \frac{w_M S^2\left(L + \frac{T_M}{S}\right) + w_H A_H}{w_M S^2 + w_H}$$

In (4), T_M/D is the slope that would result if the rise has its target magnitude T_M, given a rise duration of D. So (4) states that the actual slope of the rise is a weighted average of the slope that would yield the target magnitude and the slope target, T_S, where the weighting of T_S decreases as D^2 increases. The magnitude can then be derived from this expression by multiplying it by duration, D.

(4) $$S = \frac{w_M D^2 \frac{T_M}{D} + w_S T_S}{w_M D^2 + w_S}, \quad M = DS$$

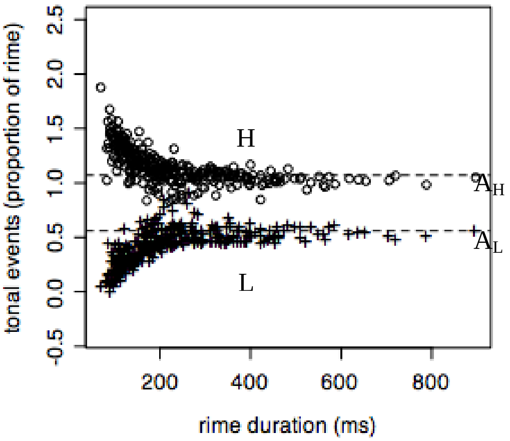
The model parameters were estimated by fitting these relationships to the data using the *R* function *nls* for non-linear least squares regression. Only the ratios of the constraint weights are relevant, so one constraint can be fixed: w_M was set to 1. The model of *M* in (4) was fitted rather than the model of *S* since error variance was more uniform for *M*. This model did not converge if all parameters were estimated simultaneously, but T_M also appears in (2) and (3), so the value of T_M was estimated based on the fits of these models. The models for *L* and *H* yielded slightly different estimates for T_M, 62 Hz from (1) vs. 68 Hz from (2), but given their standard errors, these values were not significantly different ($t(310)=1.3$, $p=0.17$). The average of these two estimates, T_M = 65 Hz, was adopted as the best estimate for all the models. The resulting parameter values are: $w_M=1$, $w_L = 0.24$, $w_H = 0.25$, $w_S = 19030$, T_M = 65 Hz, T_S = 0.48 Hz/ms. A_L and A_H were also re-estimated using models (2) and (3), but were barely changed.

The fitted models of the relationships between *M* and *D* and *S* and *D* are plotted over the relevant data in figs. 4 and 5. Fig. 6 shows fitted values of *L* and *H*, given the observed slope and timing of the other tone. It can be seen that the model accounts for all of the qualitative patterns observed in section 3: As rise duration increases, rise magnitude increases rapidly at first then levels off, and slope decreases. *L* and *H* are progressively further apart than their anchors at lower rime durations. The relationships in (2)-(4) also provide reasonable quantitative fits to the data: the residual standard errors for the models of *L* and *H* are 31 ms and 28 ms, respectively, while the error for the model of *M* is 17 Hz.

5. CONCLUSIONS

The Mandarin rising tone has a target for the slope of the rise, contrary to the SAH, but it also has targets for segmental alignment of rise onset and offset and for the magnitude of the rise, as in the SAH. These targets conflict, so the actual realization of tone is a compromise between them that depends on the duration of the syllable with which the tone is associated. This analysis is given a quantitative formalization in terms of a model of phonetic grammar based on weighted, violable constraints.

Figure 6: Modeled values of H (circles) and L (+) as proportions of rime duration plotted against rime duration.



6. REFERENCES

[1] Arvaniti, A., Ladd, D.R., Mennen, I. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26, 3-25.

[2] Chen, Y., Gussenhoven, C. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics* 36, 724-746.

[3] Flemming, E. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18, 7-44.

[4] Goldsmith, J.A. 1976. *Autosegmental Phonology*. Ph.D. thesis, MIT.

[5] Kochanski, G., Shih, C., Jing, H. 2003. Quantitative measurement of prosodic strength in Mandarin. *Speech Communication* 41, 625-645.

[6] Ladd, D.R. 2004. Segmental anchoring of pitch movements: Autosegmental phonology or speech production? In Quene, H., van Heuven, V. (eds.), *On Speech and Language: Essays for Sieb B. Nooteboom*, Utrecht: LUT, 123-131.

[7] Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation.* Ph.D. thesis, MIT.

[8] Welby, P. 2006. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34, 343-371.

[9] Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179-203.

[10] Xu, Y., Wang, E.Q. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33, 319-337.