

# The tonal timing and shape of the English L+H\* pitch accent\*

Hyesun Cho  
(Seoul National University)

**Cho, Hyesun. 2011. The tonal timing and shape of the English L+H\* pitch accent.** *Studies in Phonetics, Phonology and Morphology* 17.2, 283-311. This paper examines the timing of tonal targets in the English L+H\* pitch accent. It has been shown that the tonal targets in English are aligned with respect to certain segmental landmarks (Ladd et al. 1999, Dilley et al. 2005). The effects of segmental anchoring are re-examined in this paper. We replicate the basic effect, but also find that the tones systematically deviate from their alignment points, so that the rise starts earlier and terminates later under time pressure. Considering this to be evidence for a target duration, we propose that both segmental alignment and target duration simultaneously affect the timing of L and H tones that comprise the L+H\* pitch accent in English. The proposed model takes these two factors as weighted constraints (Flemming 2001). In this model, the actual timing of tonal targets is determined by the interaction of the alignment and duration constraints. The model parameters, the constraint weights and target duration, are obtained by fitting the proposed model to the actual data. The extended model with slope, magnitude, and alignment targets (Cho and Flemming 2011) is also applied to our experimental data. The results in this paper suggest that the phonetic realization grammar of the English L+H\* pitch accent must specify for not only the alignment of tonal targets but also the target slope and magnitude of the rise. (Seoul National University)

Keywords: tonal timing, F0, English, pitch accent, weighted constraints

## 1. Introduction

This paper examines the timing of the L and H tonal targets comprising the English L+H\* pitch accent. In Autosegmental Phonology, rising or falling contour tones are analyzed as a sequence of L and H level tones (Goldsmith 1976). In standard analyses of the phonetic implementation of tone sequences (e.g. Pierrehumbert 1980), the actual F0 contours are considered to be interpolations between level L and H targets. There is no target for the transition itself, rather the transition is considered to be a property that is derived from the scaling and location of L and H tonal targets. The research on tonal timing investigates how these tonal targets are temporally coordinated with the segmental structure.

The complication is that the pitch events corresponding to the tones (F0 peaks and valleys) do not necessarily occur within the syllable where the

---

\* I am very thankful to Edward Flemming for his invaluable advice in all aspects of writing this paper. I also thank Michael Kenstowicz for his helpful advice during the initial stage of this research. I thank the three anonymous reviewers for their helpful comments.

tones are phonologically associated. For example, according to ToBI (Silverman et al. 1992), the L+H\* pitch accent is characterized by H aligned with the prominent syllable (the stressed syllable in the prominent word), and the F0 rises through the prominent syllable. However, when the English L+H\* pitch accent is realized on the stressed second syllable in *amenable*, the F0 peak often occurs in the postaccentual vowel (Ladd et al. 1999:1550, Table 5), much after the end of the stressed syllable *-me-*. In some languages, the actual occurrence of the F0 peak may be delayed even farther than one syllable under time pressure, if the tone-segment association is weak, i.e. if the phonological association of the tone is not lexically specified but assigned at the phrasal level. Such a phenomenon is found in Seoul Korean (Cho 2011). It is proposed that such deviation of the tonal targets is to avoid too short or too long rises under variations in segmental duration due to changes in speech rate. In other words, not only the alignment between tones and segments, but also the duration between L and H targets is regulated by a constraint. On the other hand, it has been shown that English has a rather strong segmental alignment pattern, as will be reviewed in this section (Ladd et al. 1999, Dilley et al. 2005). This paper aims to examine whether a tendency to preserve a certain target duration for rises exists even in a language like English. To examine the alignment pattern of English, we manipulate speech rate, using a similar method to the Korean experiment (Cho 2011). Based on the experimental results, we argue that the L+H\* pitch accent must be specified not only for the alignment of H\*, but also for shape targets, such as duration, slope, and magnitude of the rise. Thus, constraints for the shape of F0 movements must be present in both languages.

A tendency has been noted in many languages that tonal targets are rather stably aligned with respect to certain segmental landmarks. For example, Arvaniti et al. (1998) varied the duration of accented syllables carrying a prenuclear pitch accent in Greek: [pa'remvasi] (long), [ro'ditiko] (short). They measured the distance between the L and H targets and the distance between the beginning of the postaccentual vowel and the H. The H peak was found on average 17 ms into the postaccentual vowel and this timing was not significantly affected by variations in the syllable duration. The L was also consistently aligned to the beginning of the accented syllable. On the other hand, the interval between the L and H targets was significantly affected by the duration of the accented syllables. This means that the location of the L and H tones closely follows certain segmental landmarks (the beginning of the accented syllable for L, the beginning of the vowel for H). As a result, the duration and the slope of the transition from L to H vary depending on the duration of the accented syllable.

In the case of English, it has also been shown that the tonal targets are stably aligned with respect to segments. Ladd et al. (1999) examined the timing of the beginning and end points of prenuclear accentual rises in British English. They varied the duration of syllables by changing speech

rate. They examined several segmental landmarks, looking for a segmental point whose distance to a tonal target is not affected by variation in segmental duration due to variation in speech rate. Such points existed, and the best alignment point can be expressed in terms of some proportion into the syllable, rather than a fixed segmental landmark such as the end of the vowel, etc. They calculated the alignment of H as a proportion of the duration of the consonant following the stressed vowel. A value of 1 indicates alignment at the onset of the vowel following the stressed syllable. Overall, the ratios were 1.05 in fast speech, 1.09 in normal speech, and 1.44 in slow speech, with speaker variation. In normal speech, the ratio ranged from 0.42-1.84 depending on speaker. Based on these results, Ladd et al. (1999) proposed the 'segmental anchoring hypothesis', which states that the beginning and the end points of rising F0 movements are aligned stably ('anchored') with respect to certain segmental landmarks ('anchors'). They rejected the hypothesis that duration of a rise is constant, and argued that pitch movements should be viewed as the result of interpolation between pitch targets which are aligned with respect to segments. If pitch targets are aligned with respect to segment, the duration of the pitch rise must change when the segment durations change due to speech rate or syllable structure. Thus, the segmental anchoring hypothesis is at odds with the hypotheses referred to as the 'constant duration' or 'constant interval' hypothesis (Ladd et al. 1999:1552, Dilley et al. 2005:117), which states that the rise duration is relatively stable.

Dilley et al. (2005) also examined which hypothesis is correct in English, the segmental anchoring hypothesis or the constant interval hypothesis. They manipulated the timing of L in English L+H\* pitch accents by varying the location of the word boundary, early (*Norma Nelson*) and late (*Norman Elson*). It has been observed that when each word in such a phrase carries H\*, there is an F0 minimum between the two H\*s, and the F0 minimum aligns to the word boundary between the two words (Ladd and Schepman 2003). Using the same speech materials, Dilley et al. (2005) found that the timing of H\* on the second word was not affected by variations in the timing of the preceding L. On the other hand, the interval between the L and H tonal targets was significantly affected by the timing of L. This means that L and H are aligned independently of each other, which is consistent with the segmental anchoring hypothesis.

Thus, the previous studies support the segmental anchoring of tones in English, over the constant duration or constant interval hypothesis. This paper re-examines the segmental anchoring effect in English with speech rate varied. We replicated the segmental anchoring effect in our data. At the same time, we found some tendency toward a constant duration despite the conflicting nature of the two hypotheses. A modest but systematic pattern of deviation of tones from their anchors was observed. This is interpreted to mean that there is a tendency to maintain a certain target duration between the L and H tones. That is, instead of strictly following

the anchoring points, the tones slightly deviate from their anchors to avoid too short or long a rise duration. Given that duration is derivable from slope and magnitude, we go on to explore the possibility that this tendency towards constant rise duration actually serves to maintain slope and magnitude targets of the rise. The results are modeled using weighted constraints for scalar phonetic representations (Flemming 2001).

## 2. The experiment

### 2.1 Experimental methods

The experiment aims to examine the tonal timing patterns in English. In particular, we look for the effects of segmental anchoring in English, i.e. whether the tonal alignment is stable despite the changes in syllable duration brought about by speech rate manipulations. Note that rate manipulation is to obtain a wide range of syllable duration, rather than out of a pure interest in speech rate itself. The primary purpose of varying syllable duration is to vary time pressure on the production of a given tonal melody.

At the same time, we can also compare our results of the effects of speech rate on tonal alignment in English with the results of previous studies. As mentioned in Section 1, Ladd et al. (1999) found in British English that there is a segmental landmark where a tone is stably aligned under changes in speech rate, which gives rise to the segmental anchoring hypothesis. However, the effect of speech rate on peak timing in English has been reported in Silverman and Pierrehumbert (1990). In Spanish, pitch peaks tend to occur later with respect to a segmental landmark in fast speech than in slower speech (Prieto and Torreira 2007).

It is a matter of debate whether L+H\* is categorically distinct from H\*. Pierrehumbert (1980) argues that the two are different categories, but Ladd and Schepman (2003) argues that in both H\* and L+H\* there are distinct L and H targets, with the L realized differently, so the two should be considered as a single accent category. According to Ladd and Schepman, the difference between H\* and L+H\* is the degree of emphasis, so the L+H\* tones are at the emphatic end of this continuum.

We elicited the L+H\* accent, but it is not crucial to our analysis whether it is categorically distinct from an H\* accent. We only needed a context that consistently elicits L+H\* (or the emphatic version of H\*). According to the previous literature, correction is such a context. It is argued that the meaning of the H\* accent is rather neutral (a plain response to a question like ‘Who made the marmalade?’ – ‘Maria(H\*)nna made the marmalade’, Figure 1(a)), whereas the L+H\* conveys contrastive focus (‘Bob made the marmalade’ – ‘(No.) Maria(L+H\*)nna made the marmalade’, Figure 1(b)) (Brugos et al. 2006). The rise starts near the beginning of the stressed syllable in L+H\*, whereas the rise for H\* starts from the phrase onset

(Brugos et al. 2006, Ch.2.5 p.5). The rise for L+H\* is steeper than the rise for H\*, and the beginning of the rise is much lower in L+H\* than in H\*.

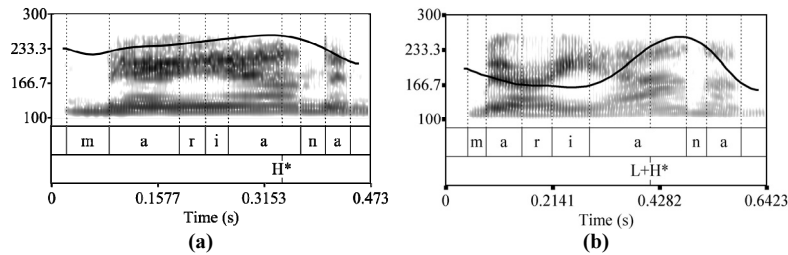


Figure 1. Examples of (a) H\* vs. (b) L+H\* (reproduced from the corresponding sound files in Brugos et al. (2006))

Thus, for our experiment, we intended to elicit the L+H\* by using a carrier phrase, 'No, I meant \_\_\_\_' – the correcting context. We had 19 target phrases and 9 fillers. The target phrases consisted of two or more words, combinations with various parts of speech, e.g. *amenable meanings*, *malaria in Nigeria* (See the Appendix for the complete list). The target melody was an L+H\* pitch accent on the stressed second syllable in the first word in the phrase, e.g. the pitch accent on the second syllable in *a'menable* in *a'menable 'meanings*. Only the words with stress on the second syllable were used in order to identify the L targets more easily by allowing enough time before the stressed syllable. Subjects were asked to read the sentences as if they were correcting the first word of each phrase, which was italicized on the list presented to the speakers. The context was straightforward and unambiguous, so speakers produced most of the target words as intended.

The stressed syllable carrying the pitch accent was followed by at least two unstressed syllables to avoid tonal crowding effects. The stressed syllables and the ones before and after the stressed syllables consisted entirely of sonorant segments to avoid pitch tracking errors near the pitch accent of interest (except one case, *Nor'wegian marinas*. This was replaced with *re'maining minutes* for the two speakers recorded later).

The filler phrases consisting of two monosyllabic words (e.g. *tip toes*) were included in order to prevent fixed rhythm. The sentences were randomized and, for the same purpose, rearranged so that no more than three target phrases came in succession. The list of sentences was presented to the speakers on a sheet of paper.

Four native speakers of English, two female (E1, E2) and two male (E3, E4), were recorded reading the speech materials. One female speaker (E2) was a native speaker of English from Canada, and the other three speakers were native speakers of American English. The speakers were naive to the purpose of the experiment. The speakers were asked to read the speech materials at normal, fast, and then slow speech rates. The rates were self-

selected. At first, they were asked to read the materials naturally, with no instructions concerning speech rate. Then they were asked to read as fast as possible, and then to read slowly but naturally. The whole list was read twice at each rate. Thus, the total number of target tokens was  $19 \text{ (phrases)} \times 4 \text{ (speakers)} \times 3 \text{ (rates)} \times 2 \text{ (repetitions)} = 456$ . The recordings were made in a sound-attenuated recording booth in the phonetics lab at the MIT Linguistics Department.

## 2.2 Measurement

F0 minima, F0 maxima, and segmental boundaries were manually marked using Praat (Boersma and Weenink 2010). Additional landmarks in the F0 trajectory, the ‘inflection points’ or the ‘elbows’, were identified by approximating the trajectory with three straight lines (adaptation of the two-piece linear regression described in Welby (2006)). The inflection points are the points where the slope of the F0 contour abruptly changes. The English pitch accent is often preceded by a plateau, so in such cases there is no clear F0 minimum (e.g. Figure 3). The minimum may reflect unimportant fluctuations in the low plateau preceding the beginning of the fast rise.

Applying the three-piece linear regression, the inflection points correspond to the two intersection points of the three fitted lines (See Figure 2). Figure 2 shows the actual fitted F0 curves using three-line fitting for (a) concave, (b) sigmoid, and (c) convex shapes. In each panel, the small circles represent the actual F0 curve. An F0 curve is divided into three parts, using three lines fitted over the actual data points. Two vertical lines (one solid, one dashed) correspond to the intersections of the three fitted lines. These vertical lines indicate the location of the elbows (=inflection points) in each F0 curve. The solid line is the location of the first elbow, and the dashed line is the location of the second elbow.

The shape of an F0 curve is determined automatically using the relationships between slopes of the three fitted lines. If the last line segment is the steepest, it is a concave shape. If the second line segment is the steepest, it is a sigmoid shape. If the first line segment is the steepest, it is a convex shape.

In this way, there are two inflection points, or two elbows, for each rise. Of the two, we take the one at the beginning of the fast rise to be the location of the L target. That is, for concave rises, the second elbow is the L. For sigmoid rises, the first elbow is the L. The F0 minimum is the L only for convex rises, but these are very rare. In sum, in this paper we will use the L targets determined by these procedures, which corresponds to the beginning of the steepest line segment in the three-piece linear regression. For comparison, we will later provide the analysis where the F0 minima are taken to be the L. The H was always taken to be the F0 maximum of the rise.

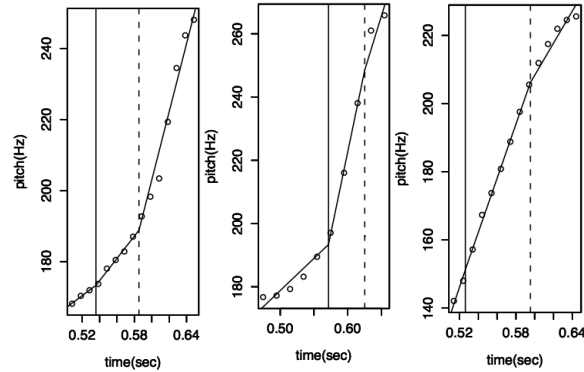


Figure 2. F0 curves fitted using three-piece linear regression. (a) Concave (*Millennium Resort*, E2), (b) Sigmoid (*eliminate Mariners*, E2), and (c) Convex (*anonymous writing*, E4)

### 3. The results

#### 3.1 Overall shape of the L+H\* pitch accent

Figure 3 shows examples of the elicited rises. As can be seen, the rises were similar to the L+H\* accent shown in Figure 1(b), rather than H\* in that F0 remains low until the beginning of the stressed vowel, followed by a steep rise.

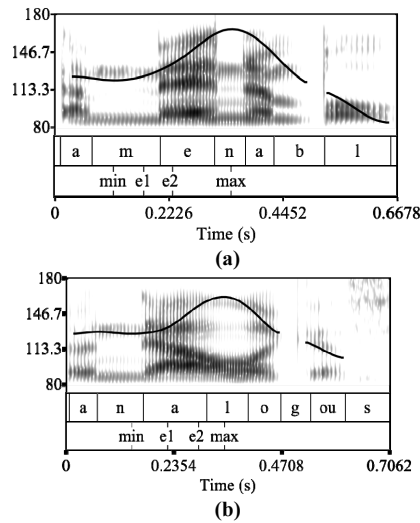
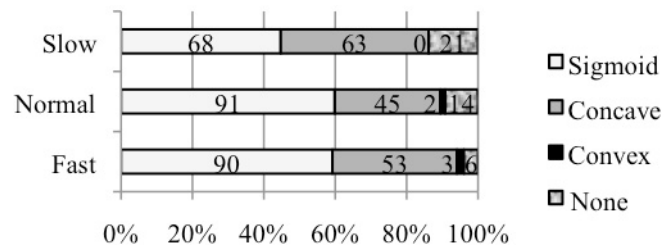


Figure 3. Examples of the L+H\* pitch accent (sampled from our recording), speaker E3's pitch tracks of (a) *amenable*, (b) *analogous*. 'min' F0 minimum, 'max' F0 maximum, 'e1' the first elbow, 'e2' the second elbow.

The shape classification can be additional evidence that the majority of accents were L+H\*'s. The majority of English rising pitch accents recorded were sigmoid or concave in shape (55%, 35% respectively, across all speech rates), as shown in Table 1. In fast and normal speech, sigmoid rises are almost twice as frequent as concave rises. Convex shapes are very rare in general, accounting for only 1% of the whole data. The label 'none' refers to shapes that are not classifiable. These are usually due to segmental perturbation along the rise, so they were discarded in the analysis.

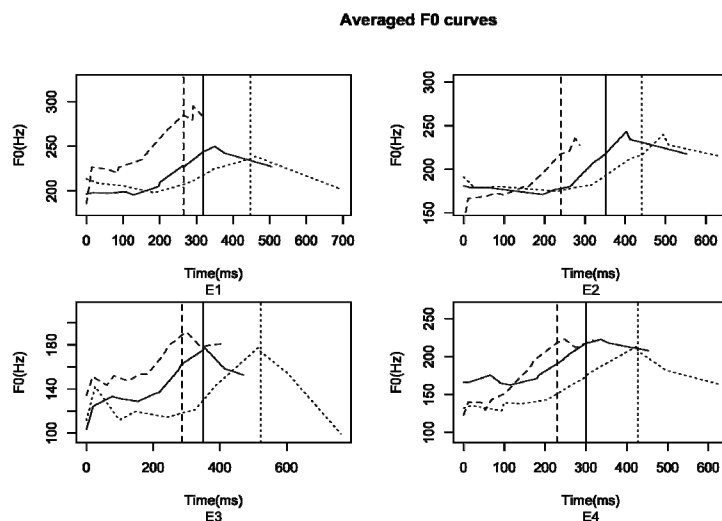
Table 1. The shape of rises



Although we intended to elicit the bitonal L+H\*, we cannot exclude the possibility that there may have been a few H\* pitch accents, probably the ones classified as convex shapes which accounted for only 1% of the total tokens. Convex shapes start with a fast rise, so it is possible that the convex shapes may be H\* accents, but in any case, their frequency is extremely low. A closer examination reveals that even these convex rises have significantly low L targets (e.g. Figure 2(c)), so they still look more like L+H\*.

Furthermore, the average shapes are closer to L+H\*. Figure 4 shows the averaged F0 curves of the portion from the beginning of the words up to the beginning of the vowel in the third syllable, for each speaker. Vertical lines show the end of the stressed second syllable. The averaged shapes show that the F0 remains low until the fast rise which terminates near the end of the stressed syllable (cf. Figure 1); this means that the elicited pitch accent is L+H\* rather than H\*. For most speakers, the average peak of the rise is found at the end of the second syllable in all speech rates. The peaks of the rises of speaker E2 occur later than the end of the second syllable, but the timing relative to the end of the syllable seems fairly consistent across speech rates within the same speaker.





**Figure 4.** Averaged F0 curves for each speaker. Solid lines: normal speech, dashed lines: fast speech, dotted lines: slow speech. The vertical lines correspond to the end of the stressed second syllable.

### 3.2 Segmental anchoring

If the L and H tonal targets are aligned with respect to certain segmental anchors, then there should be a positive linear relationship between the timing of a tone and its anchor. This section examines whether such a linear relationship is found.

To begin, we need to find where the segmental anchor is. Several segmental landmarks were tested to identify the segmental landmark that has the highest correlation with L or H tonal targets. We examined the correlation between the time of the L or H targets with the time of each candidate segmental landmark. The times were all measured from the beginning of the target phrase. The beginning is time 0, and the times of segmental landmarks, L or H targets are the duration in milliseconds from this phrase onset. The tested landmarks are various segmentally-defined positions in the first two syllables, as follows: In the first syllable: the beginning of the vowel ('v1'), the middle of the vowel ('vm1'), the middle of the rime ('rm1'); in the second syllable: the beginning of the onset consonant ('c2'), the beginning of the vowel ('v2'), the middle of the vowel ('vm2'), and the middle of the rime ('rm2'), and the end of the second syllable ('f2').

We searched for the best-correlated point to approximate the anchor for the given tone. Linear regression models were fitted to the data with the timing of a tone as the dependent variable, and the timing of a segmental

position, speaker and their interaction as predictor variables. The correlations between a tone and the candidate segmental anchors listed above were compared one by one. The best correlation for L was found with the middle of the second rime ('rm2') ( $R^2=0.87$ ). The best correlation for H was also found with the middle of the second rime ('rm2') ( $R^2=0.93$ ).

Since these results are pooling across speakers, we also tried fitting linear mixed-models (LME) where speaker-dependent variations are treated as random effects rather than fixed effects. The LME is more appropriate to treat speakers as drawn at random from a larger population. The timing of L was the dependent variable, the timing of a segmental position was a fixed effect, and by-speaker random intercepts and slopes for the timing of a segmental position were included in the model specifications. Deviance is the basic measure of goodness of fit in linear mixed-effects models: the lower the deviance, the better fit. For L, the mixed model with the segmental position 'rm2' had the lowest deviance (3595) among other points. For H, 'rm2' had the lowest deviance (3736). Based on these models, the tentative anchoring points are taken to be 'rm2' for both L and H. The result that the best anchoring point was the same for L and H may suggest that the L and H tones comprising the pitch accent are closely related, as a unit (cf. Ladd 2004b). However, these are just first approximations of the anchors, and precise locations of anchors will be re-estimated in later sections (Section 4.2.2 and 5.2).

Figure 5 plots the timing of the L and H against their tentative anchors. The solid line is the regression line between the anchor and the given tone. The dashed line is the  $y=x$  line, i.e. the line that is expected if the tone falls on the anchor exactly. Or if the anchor is in fact the middle of the second rime *plus* some fixed milliseconds (as in the case of Greek: H peaks occurred 17 ms after the onset of the post-accentual vowel), then the regression line should be parallel to the  $y=x$  line (i.e. the slope of 1). The  $R^2$  values under the plots are pooling across all speakers, so they are different from the  $R^2$  values used in the selective search mentioned above, where speaker is treated as a fixed effect.

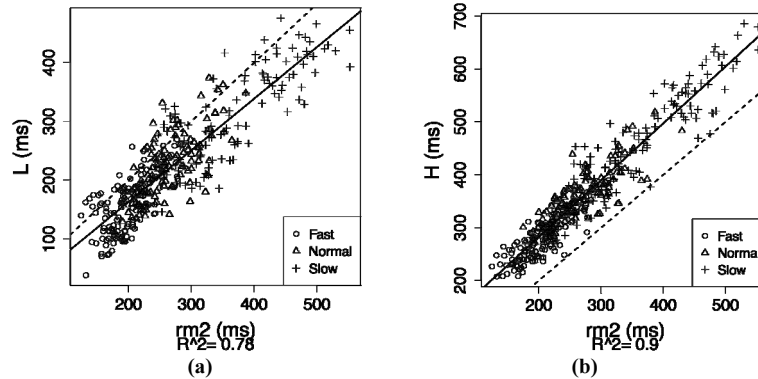


Figure 5. Alignment of L (elbows) and H (F0 maxima): (a) L against  $A_L$ , (b) H against  $A_H$ . The dashed line is the  $y=x$  line. The anchor 'rm2' is a tentative approximation of the anchor.

In both L and H, a tone and its anchor show a positive linear relationship. This is as predicted by the segmental anchoring hypothesis, so we can say that segmental anchoring is observed in our data. For L, the slope of the regression line was significantly different from 1 (slope 0.87,  $t(366)=5$ ,  $p<0.0001$ ). For H, the slope of the regression line was slightly different from 1 (slope 1.07,  $t(388)=3.8$ ,  $p<0.001$ ).

The purpose of manipulating speech rate was to have a range of variation in segmental duration so that the L and H tones could be realized in varying time pressure. In this sense, speech rate manipulation was successful. Considering the distribution of rime duration ('rm2') along the abscissa in Figure 5, the distribution was fairly continuous, with a bit wider range in slow rates (the reason may be that more variation is possible in slow rates whereas there is a limit on how fast one can speak).

### 3.3 Tendency toward a target duration

Along with the effects of segmental anchoring, systematic deviations from the anchoring point were observed, as a function of speech rate. Figure 6 plots the deviation of the given tone from its anchor, normalized by speech rate (the inverse of the duration of the first two syllables). The dashed line is the location of the anchor. The data points indicate the relative location of the given tone. In Figure 6(a), the L's were found gradually earlier relative to the anchoring point in faster speech, and later in slower speech. The tendency was very small, but significant (slope=-0.84,  $t(281)=-2.27$ ,  $p<0.05$ ). In Figure 6(b), the H's were found closer to the anchor in slow speech, later in fast speech. The slope was significantly different from zero (slope=4.77,  $t(388)=13.9$ ,  $p<0.0001$ ). The substantial delay in the H peak with respect to the anchor translates into delay of larger amounts of

segmental material at faster speech rates. That is, at faster rates, segments are short, so relative delay becomes greater. In sum, the pattern is that the rise starts earlier and terminates later when there are more time pressure due to increased speech rate.

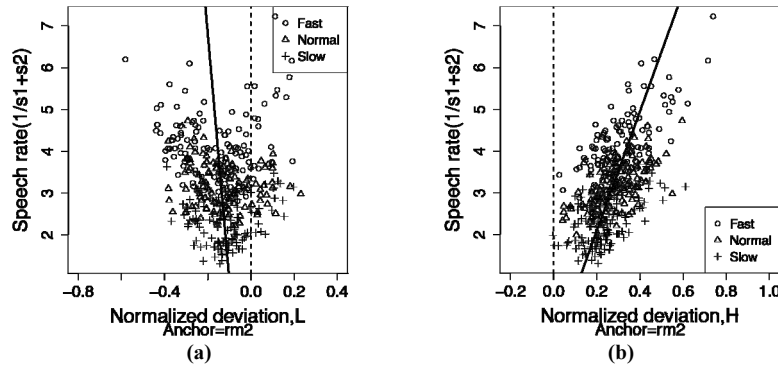


Figure 6. Deviation of the L and H from the anchor. The anchor is the middle of the second rime ('rm2'). (a) L deviation from  $A_L$ , ( $R^2=0.014$ ) (b) H deviation from  $A_H$ . The dashed line is the position of the anchor ('rm2') ( $R^2=0.33$ ).

We interpret these results to mean that there is a *target duration* that regulates the distance between L and H. That is, there is a tendency for the L and H tones to follow their anchor, but the tones deviate from their anchors in order to avoid too short a rise (in fast speech) or too long a rise (in slow speech). This pattern is similar to the tonal alignment patterns shown in Prieto and Torreira (2007) for Spanish. In Spanish, peaks tended to occur later than a segmental anchor in fast speech. They only showed the results categorically for three speech rates: fast, normal, and slow. On the other hand, Figure 6 shows gradual effects of speech rate along a continuum of local speech rates.

To sum up the experimental results, we found a tendency toward segmental anchoring, along with a tendency to maintain a target rise duration. Conceptually, satisfying segmental alignment and target rise duration both might have perceptual and articulatory motivations. Segmental alignment is important to signal the phonological association between a tone and a syllable, so in many languages segmental alignment is found to be relatively stable (e.g. Arvaniti et al 1998, Ladd et al. 1999, Dille et al 2005). Maintaining a certain target duration may also be important due to perceptual or articulatory reasons. It might be supposed that the duration target has a basis in ease of articulation considerations. It is well known that there are physiological limits on the time required to produce a change in pitch (Sundberg 1979, Xu and Sun 2002). However, these articulatory considerations would motivate a constraint requiring only a minimum duration for the rise, whereas in the experiment we also

observed avoidance of rises that exceed the target duration, i.e. the L is shifted later and the H is shifted earlier toward their anchor in slower speech. Thus, perceptual motivations might be more appropriate to explain both directions of deviation. An overly short rise would be hard to detect (Greenberg and Zee 1977, Thyer and Mahar 2006), and an overly long rise would be too shallow to sound like a rise. At this point, we cannot be certain that the proposed target duration is directly determined by either articulation or perception. Rather, we suppose that it is a linguistically determined target within the range that articulation and perceptual capacities allow.

### 3.4 F0 minima

For the reasons stated in Section 2.2, we used elbows as the location of the L tonal targets. That is, the L+H\* accent is often preceded by a low plateau, so the F0 minima measured in the plateau are not meaningful in the F0 contour of the pitch accent. In addition, in Section 3.2 and 3.3, we showed that the elbows are stably aligned with a segmental landmark, which supports our usage of elbows as L tonal targets. In this section, we examine the F0 minima to compare them with the elbows we have used for the L tonal targets.

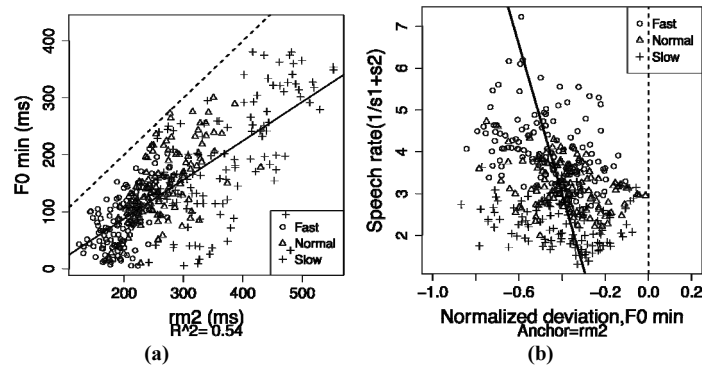


Figure 7. (a) F0 minima against 'rm2'. The dashed line is the  $y=x$  line ( $R^2=0.54$ ). (b) F0 minima deviation from  $A_L$ . The dashed line is the position of the anchor ('rm2') ( $R^2=0.1$ ).

Applying the same procedure as before, the middle of the second rime 'rm2' was the best-correlated point with the F0 minima ( $R^2=0.674$ ) ('vm2' was only slightly higher,  $R^2=0.676$ ). 'rm2' had the lowest deviance (4144) ('vm2' had a higher deviance of 4400). Figure 7 plots the F0 minima with 'rm2', showing the F0 minima show a more scattered pattern than the elbows. Comparing the deviation plots (Figure 7(b) and Figure 6(a)), the elbows occur closer to the anchor than the F0 minima, though the deviation pattern is similar (occurring earlier under more time pressure). To sum up,

the elbows are more stably aligned with respect to the hypothesized anchor. Thus, we continue to take the elbows to be the L targets.

#### 4. Modeling

##### 4.1 The Alignment-Duration model

As shown in the experimental results, tendencies to maintain both segmental anchoring and target duration were observed. Thus, the two are not inviolable principles determining the timing of tones, rather they should be interpreted as violable constraints. In the model we propose in this section, the actual timing is determined by a compromise between them. We model these findings using weighted constraints (Flemming 2001), following Cho (2011). The constraints and the cost of violation of each constraint are shown in (1). The alignment constraint  $T(H)=A_H$  requires that the H peak occur at the anchor. The cost of violating this constraint is the squared deviation of H from the anchor, multiplied by the weight  $w_A$ . The same applies to  $T(L)$  and  $A_L$ . On the other hand, the duration constraint requires that the duration of the rise correspond to a target duration  $T_D$  (a positive constant).

(1) Constraints and their cost of violation

|          | Constraint      | Cost of violation      |
|----------|-----------------|------------------------|
| Align(L) | $T(L)=A_L$      | $w_L(T(L)-A_L)^2$      |
| Align(H) | $T(H)=A_H$      | $w_H(T(H)-A_H)^2$      |
| Duration | $T(H)-T(L)=T_D$ | $w_D(T(H)-T(L)-T_D)^2$ |

The relative importance of these constraints is reflected in their weights ( $w_L$ ,  $w_H$ ,  $w_D$ ). The higher weight, the more important it is to satisfy that constraint. For example, if the alignment of L is more important than the alignment of H in a language, the weight of Align(L) is higher than the weight of Align(H) in that language. So, in such a language, the alignment of L will be less affected by changes in speech conditions such as speech rate. Note that Align(L) and Align(H) do not conflict with each other; satisfying both at the same is not contradictory. In other words, alignment of L does not imply violation of alignment of H. Instead, the alignment constraints conflict with the duration constraint, because satisfying alignment constraints results in violation of the duration constraint, and *vice versa*. The alignment constraint will be violated to the degree that the duration constraint is satisfied. This concept can be made quantitatively precise from the following formulation.

In the proposed model, the actual timing of L and H are determined as the values that minimize the summed cost of violations, instead of counting violation marks as in Optimality Theory (Prince and Smolensky 1993/2004). The cost function (summed cost of violations) is in (2) (cf.

Flemming 2001:20).

$$(2) \text{ Cost} = w_L(T(L) - A_L)^2 + w_H(T(H) - A_H)^2 + w_D(T(H) - T(L) - T_D)^2$$

The minimum of the cost function is found where its derivative is zero. By differentiating the cost function in (2) for  $T(L)$  and  $T(H)$  respectively, and setting each derivative function equal to zero, we obtain the expressions for the values of  $T(L)$  and  $T(H)$  where the partial derivatives are zero, as shown in (3).

$$(3) \text{ a. } T(L) = \frac{w_L}{w_L + w_D} A_L + \frac{w_L}{w_L + w_D} (T(H) - T_D)$$

$$\text{b. } T(H) = \frac{w_H}{w_H + w_D} A_H + \frac{w_H}{w_H + w_D} (T(L) + T_D)$$

(3a) implies that the actual timing of L is a weighted average of the timing of the anchor for L ( $=A_L$ ) and the timing that satisfies the target duration  $T_D$  ( $=T(H) - T_D$ ). The same holds for (3b). The actual timing of H is a weighted average of the timing of the anchor for H ( $=A_H$ ) and the timing that satisfies the target duration  $T_D$  ( $=T(L) + T_D$ ). The timing of L and H thus deviate from their alignment targets in order to partially accommodate the duration target, in accordance with relative weights of the constraints in a given language. The proposed model suggests that the actual timing of a tone is determined through trade-offs between alignment and duration constraints, resulting in deviation from its anchor depending on the relative importance how important duration constraint is compared to alignment constraint.

## 4.2 Obtaining model parameters

### 4.2.1 Constraint weights

The expression in (3a) means that  $T(L)$  is a linear function of  $A_L$  and  $T(H)$ , so we fitted linear mixed-effects models to the experimental data to find the actual model parameters, i.e. the constraint weights and the precise anchoring points. The *R* package *lme4* for linear mixed-effects models was used for model fitting. A similar procedure was applied to  $T(H)$  in (3b). To obtain unique values for the weights, we set  $w_H + w_L + w_D = 1$ , because only the ratio between the weights matters. With this condition and the coefficients of the fitted mixed models, the constraint weights were computed as  $w_L = 0.15$ ,  $w_H = 0.49$ ,  $w_D = 0.36$ . These constraint weights mean that alignment of H peaks is relatively stable in English, but also that the target duration is important.

## 4.2.2 Estimating precise anchors

The precise location of the anchors ( $A_L$ ,  $A_H$ ) can also be computed using the expressions in (3). The previous estimates of the anchors were the middle of the second rime for both L and H (Section 3.2). These estimates were based on the best correlation between a tone and a segmental landmark. On the other hand, the expressions in (3), which is obtained from the proposed Alignment-Duration model, include a duration term, in addition to the alignment term. So, the best estimates of the anchors might differ, given this model. Another difference is that the previous anchor estimates were based on a selective search – that is, several pre-determined candidate anchors were compared for the best correlation or the lowest deviance.

Now, using the AD model, the precise anchor positions can be directly computed, rather than searched for. The procedure is as follows. In the expressions in (3), we replace  $A_L$  with  $v2+p \cdot rime2$ , where  $v2$  is the beginning of the second vowel,  $rime2$  is the duration of the second rime, and  $p$  is the proportion into the rime. This reformulation allows us to express the anchor location as ‘some proportion ( $p$ ) into the second rime’. We know  $v2$  and  $rime2$  values from our experimental data. The unknown value  $p$  can be computed from the coefficients of the LME fitting described above. From here, we obtain  $p = -0.16$ . The negative  $p$  value is problematic because it means that the optimal anchor for L precedes the second rime. Thus, we refitted the model on the hypothesis that  $A_L$  is in the first syllable, rather than in the second syllable. So  $A_L$  is now replaced with  $v1+p \cdot rime1$ , where  $v1$  is the beginning of the first vowel,  $rime1$  is the duration of the first rime, and  $p$  is the proportion into the rime. From here, we obtain  $p = 0.18$ . Thus, we estimate the anchor for L ( $A_L$ ) as follows:  $A_L = v1 + 0.18 \cdot rime1$ , that is,  $A_L$  is 18% into the first rime. However, the deviance was greater in the model with  $rime1$  (3710) than in the model with  $rime2$  (3512). That is, the goodness of fit was better with  $rime2$ . We proceed with the anchor in the first rime for now, but it will be re-estimated using another model in Section 5.2. Following the same procedure,  $A_H$  was also estimated:  $A_H = v2 + 0.93 \cdot rime2$ , that is, 93% into the second rime.

Comparing our anchor estimate ( $A_H$ ) with Ladd et al. (1999:1550)’s results, our estimate is earlier. Their anchor was in the post-accentual vowel, whereas our estimate, 93% of the rime, puts the anchor in the stressed syllable. However, their speech materials were different from ours in that Ladd et al. (1999) looked at prenuclear accents and not necessarily L+H\* (e.g. ‘There was a nominal fee for his services’). It has been reported that F0 peaks are aligned earlier in nuclear accents (our condition) than in prenuclear accents (Steele (1986), Silverman and Pierrehumbert (1990) (English), Schepman et al. 2006 (Dutch), Nibert 2000 (Spanish)). Moreover, the timing is subject to dialectal variation – in American



English, Southern Californian speakers show much later alignment of L and H than Minnesotan speakers (Arvaniti and Garding 2007). Southern and Northern German also exhibit dialectal variation (Atterer and Ladd 2004). Ladd et al. (1999) looked at British English speakers, whereas our speakers were American and Canadian English speakers. Thus, the timing is not expected to be the same.

More comparable results would be found in Dilley et al. (2005) since they elicited L+H\* accents. However, they did not attempt to estimate the location of the anchor. They examined the difference between H and the onset of the postaccentual vowel only ('H-V') without claiming that this point is the anchor. They showed that the H-V measure is not affected by early-late word boundary conditions. Thus, the onset of the postaccentual vowel is a potential anchoring point. However, in our data, the onset of the postaccentual vowel had a worse fit than the points in the stressed vowel, in terms of both deviance (3865) and correlation ( $R^2=0.89$ ) (cf. the previous estimate 'rm2' had deviance of 3736 and  $R^2=0.92$ ).

#### 4.2.3 Target duration $T_D$

Using the coefficients from the fitted mixed models in 4.2.1, the target duration  $T_D$  values were estimated as 55 ms from the T(L) model, but 176 ms from the T(H) model. We will refer the  $T_D$  estimate from T(L) model as  $T_{DL}$ , and the D estimate from T(H) model as  $T_{DH}$ . It is problematic that  $T_{DL}$  and  $T_{DH}$  do not converge, because in the AD model, there is only one  $T_D$  value. We can examine whether the difference between  $T_{DL}$  and  $T_{DH}$  is significant by calculating confidence intervals for these estimates. The distribution of  $T_D$  estimates was determined through simulation: for each model, we sampled 1000 pairs of slope and intercept values based on the variances of these parameters estimated in fitting the model, and calculated  $T_D$  from each pair (following the procedure in Gelman and Hill (2007)). This gives the probability density function (PDF) of  $T_{DL}$  and  $T_{DH}$  respectively. Based on these distributions, the confidence intervals of the  $T_{DL}$  and  $T_{DH}$  values did not overlap. The 95% confidence interval of  $T_{DL}$  was 21 to 85, the 95% confidence interval of  $T_{DH}$  was 150 to 208. The difference ( $T_{DH}-T_{DL}$ ) was significantly different from zero, because the 95% confidence interval of the difference did not include 0 (79 to 166). The difference indicates a problem since the AD model posits a single value for  $T_D$ , so a solution will be discussed in the next section.

#### 4.3 The L offset from its estimated anchor

In the previous section, we have shown that the  $T_D$  values did not converge. Looking at the data, it seems that the problem lies in the timing of the L relative to the anchor. Figure 8 shows the alignment and deviation plots of L's with the  $A_L$  estimate from Section 4.2.2. In Figure 8(a), the L is plotted

against  $A_L$  (the anchor estimate in Section 4.3). In Figure 8(b), normalized deviations of  $L$  are plotted against speech rate. As the plots show, the timing of  $A_L$  is quite far from the actual occurrence of the L's. The mean of  $A_L$  was 36 ms, the mean of the  $L$  was 229 ms, and the mean of the difference between the anchor  $A_L$  and the actual occurrence of  $L$  was 194 ms.

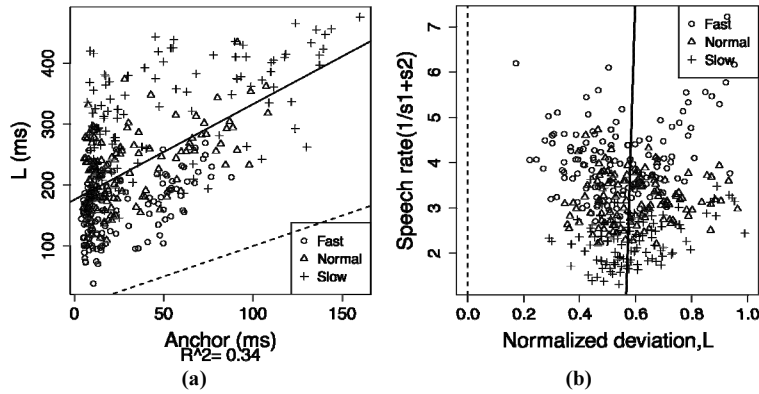


Figure 8. Alignment and deviation of the L (elbow).  $A_L = v1 + 0.18 \times \text{time1}$ . (a)  $L$  against  $A_L$ . The dashed line is the  $y=x$  line. (b)  $L$  deviation from  $A_L$ . The dashed line is the location of the anchor.

This may mean that the anchor of  $L$  is actually some offset from the estimated  $A_L$ . If so, a more accurate anchor  $A_L'$  can be expressed as  $A_L' = A_L + k$ , where  $A_L$  is the anchor estimated in Section 4.2.2, and  $k$  is an offset from  $A_L$  (a positive constant). Conceptually, this means that the English rising pitch accent starts somewhere in the low plateau (or low  $F_0$  stretch). The elbows are aligned with respect to some point during the plateau, but it is delayed by  $k$  from that point, producing a plateau before the fast rise. The duration of the plateau varies depending on the speech rate. The faster the speech rate, the shorter the delay before the fast rise. The difference between  $A_L$  and  $L$  in Figure 8(a) indicates the duration of the plateau:  $A_L$  is a point early in the plateau, and  $L$  (elbow) is the beginning of the fast rise.

For this reason, we expect that the adjustment of  $A_L$  by an offset  $k$  will improve the AD model. We replaced the alignment target  $A_L$  with  $A_L + k$  in our interpretation of the model parameters. With this, the  $T_D$  value from the H model was 176ms (the same as before). The  $T_D$  value from the L model cannot be computed directly because  $k$  is unknown. To test whether the proposed model yields a reasonable value for  $k$ ,  $T_{DH}$  is plugged into the L model as the  $T_D$  value. From this,  $k = 292$ . The confidence interval of  $k$  was 181 ms to 418 ms.

The resulting  $k$  value seems reasonable, although it has a wide confidence interval.  $k$  is the offset value of the actual  $L$  target from  $A_L$ , i.e.

it should be close to the difference between  $A_L$  and L. The mean of ( $A_L-L$ ) in the data is 194ms. By speech rate, the means are 136 ms in fast speech, 195 ms in normal speech, and 264 ms in slow speech. Given that  $w_L$  is low and it is difficult to observe the effect of a constraint that has a very low weight, the  $k$  value can be considered to be fairly close to what it is expected to be.

#### 4.4 Model comparison

Model comparison was used to see whether the duration constraint is indeed necessary in modeling tonal timing, in addition to the alignment constraints. We compare the AD model in (3) with a model without the duration term. This model is referred to as the ‘Independent Alignment’ (IA) model, where the timing of a tone is predicted by the segmental anchor only. The IA model is expressed as in (4) (a,c: the coefficients in the linear model, b,d: the intercepts in the linear model).

$$(4) \begin{aligned} T(L) &= a \cdot A_L + b \\ T(H) &= c \cdot A_H + d \end{aligned}$$

In this model, the L and H are considered to be aligned with their respective anchors, independently of each other. The timing of a tone is expressed as a linear function of its anchor only. There is no ‘duration’ target between the L and H. However, (4) is not the segmental anchoring model in its strictest sense since the intercepts and the slopes can have any values (the strict model would be  $T(L)=A_L(+b)$   $T(H)=A_H(+d)$ ) (See Cho (2011) for further discussion). Linear mixed-effects models of the forms in (4) were fitted to the data, for  $T(L)$  and  $T(H)$  respectively. The anchors were also estimated based on (4), as follows:  $A_L=v1+0.59\text{-rime1}$ ,  $A_H=v2+0.58\text{-rime2}$ . To compare the IA and AD models, we looked at AIC (Akaike’s Information Criterion) values. If the difference in AIC values for two models is greater than 10, there is ‘essentially no support’ for the model with the higher AIC (Burnham and Anderson 2002:70ff.). The AIC values of the AD model were much lower than those of the IA model: 3722 (for L), 3418 (for H) in the AD model, and 4307 (for L), 3752 (for H) in the IA model. Thus, the AD model is better than the IA model.

### 5. The model with slope, alignment, and magnitude

#### 5.1 Magnitude and slope of the rise

Until now, we have looked at the temporal, timing dimension of the English pitch accent. In particular, we have shown that a duration target significantly contributes to the timing of the tonal targets. In this section, we ask whether the rise is really subject to a duration target, or whether the

duration target could be instead due to a combination of magnitude and slope targets. The duration of a rise is derivable from the magnitude and slope, and perhaps the magnitude and slope of the rise are more intuitively salient properties of an F0 movement than its duration, but the AD model did not include these properties. It has been observed that in British English, the magnitude of the F0 rise in a rising pitch accent tends to remain relatively stable under time pressure, so when the anchors get closer to each other at a fast rate, the slope of the rise gets steeper, and the duration of the rise gets shorter (Ladd et al. 1999). On the other hand, in Russian, the magnitude of rising pitch accents is reduced at faster rates, while the tonal targets are segmentally aligned (Igarashi 2004). This implies a relatively constant slope. A similar pattern is found in French (Fougeron and Jun 1996). This subsection examines how the magnitude and slope change depending on rise duration in our data. Then we will apply a model with magnitude and slope targets (Cho and Flemming 2011) to our English data in the next subsection.

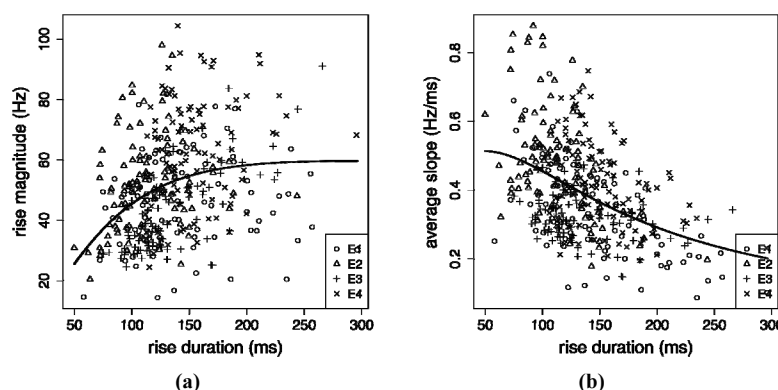


Figure 9. (a) Magnitude against rise duration (b) Slope against rise duration. The fitted curves are drawn using the relationships in (9) in Section 5.2, with estimated parameter values in (10) plugged in.

Figure 9 shows the relationship between rise duration and magnitude, and the relationship between rise duration and slope, observed in our data. Ladd et al. (1999) found that the magnitude of the rise is not affected by rise duration in four out of six speakers (Ladd et al. 1999:1552). However, in our data, rise duration had a significant effect on rise magnitude. A linear mixed-effects model was fitted to examine the effect of duration on the rise magnitude. The dependent variable was rise magnitude, the fixed effect was rise duration, and there were by-speaker random intercepts. The model with the duration fixed effect was better than the model without the duration fixed effect, according to a Likelihood Ratio Test ( $\chi^2(1)=35$ ,  $p<0.0001$ ). The effect of duration on magnitude was small (0.12) but

significant. Thus, there was a slight tendency that rise magnitude increases as rise duration increases – this means that the slope tends to stay close to a target slope, because if slope were constant, magnitude would increase with increased duration. However, the increase in magnitude is insufficient to yield a constant slope, since slope decreases as rise duration increases (Figure 9(b)).

The slope is inversely correlated with duration, as shown in Figure 9(b). This would necessarily be the case if rise magnitude were constant, given that  $\text{slope} = \text{magnitude}/\text{duration}$ , but as pointed out, magnitude increases with duration, so it is not obvious that the slope should decrease with increasing duration. This plot shows that it does. To examine the effects of duration on slope, a linear mixed-effects model was fitted to the data with slope as dependent variable, duration as a fixed effect, with by-speaker random intercepts. This model was better than the model without the duration fixed effect ( $\chi^2(1)=79.9$ ,  $p<0.0001$ ). The effect of duration on slope was significant (slope=-0.001). The effect (-0.001Hz/ms) is not small. The duration is in ms, so a change in duration of 100ms yields a change in slope of -0.1Hz (100×-0.001). If, for example, the slope was 0.4 Hz/ms at a duration of 150ms (Figure 9(b)), the magnitude would be 60Hz. Now when the duration increases to 250ms, the slope is predicted to become 0.3 (=0.4-0.1), then the magnitude would be 75Hz (=0.3×250). If the slope had not changed with duration, the magnitude would have been 100Hz (=0.4×250). In sum, slope becomes steeper as rise duration gets shorter – a comparable result to Ladd et al. (1999). However, both slope and magnitude were significantly affected by the duration, that is, neither slope nor magnitude is invariant.

One might suspect that the relationships between slope/magnitude and rise duration could be governed by physiological constraints such as ease of articulation or perception. Too steep a slope would require too much effort, and too shallow a slope would be difficult to detect. The target slope could be determined within this range. While such explanations could be reasonable, in this paper we do not intend to directly correlate the proposed targets (duration, slope, and magnitude) with physiological factors. Articulatory efforts are not always minimized in speech production. Speakers can do things sacrificing their articulatory comfort – e.g. slope becomes steeper when rise duration decreases (Figure 9(b)), while at the same time they seem to reduce articulatory efforts by reducing magnitude as rise duration decreases (Figure 9(a)). We tentatively assume that the targets are linguistically defined by language-specific grammar, within the range where articulatory and perceptual requirements allow.

## 5.2 The model with slope, magnitude, and alignment (SAM)

In Section 5.1, we have shown that magnitude and slope of the rises are affected by rise duration. Thus, we can see that all properties of the L+H\*

rise - slope, magnitude, and alignment - are subject to variation depending on time available for the production of the rise. We interpret these three properties as targets. Duration is derivable from slope and magnitude, so given slope and magnitude targets, the duration target is not necessary (but one could imagine a model with all three targets, and this possibility can also be explored).

The model in (1) in Section 4.1 can thus be revised with magnitude (M) and slope (S), as shown in (5). This model is developed for the Rising tone in Mandarin in Cho and Flemming (2011). In this section, we introduce and apply this model to the English data.

(5) Constraints and the cost of violations

|           | Constraint | Cost of violation   |
|-----------|------------|---------------------|
| Align(L)  | $L=A_L$    | $w_L(L-A_L)^2$      |
| Align(H)  | $H=A_H$    | $w_H(H-A_H)^2$      |
| Magnitude | $M=T_M$    | $w_M(S(H-L)-T_M)^2$ |
| Slope     | $S=T_S$    | $w_S(S-T_S)^2$      |

(5) shows the constraints in the model with slope, alignment, and magnitude (SAM). The alignment constraints are the same as before: the constraints require the tones to be aligned with respect to their anchors.  $T_M$  is the target magnitude of the rise. Any deviations from the target magnitude incur a cost for violating the magnitude constraint. The same holds for the slope constraint. That is,  $T_S$  is the target slope of the rise, and any deviations from the target slope incur a cost. Note that we do not have the duration constraint anymore, assuming the duration is determined by magnitude and slope. The magnitude constraint replaces M with  $S(H-L)$ , where  $(H-L)$  is the actual duration between L and H.

The actual timing of L and H tonal targets and the actual slope and magnitude values are determined by finding the values that minimize the summed cost of violations of the constraints. The cost function is shown in (6).

$$(6) \text{ Cost} = w_L(L-A_L)^2 + w_H(H-A_H)^2 + w_M(S(H-L)-T_M)^2 + w_S(S-T_S)^2$$

The minimum of this cost function is found where its derivative is equal to zero. We differentiated this cost function with respect to L, H, and S respectively, and set each partial derivative to zero. From this, we have the following relationships for L, H, and S (where  $D=H-L$ ).

$$(7) L = \frac{w_M S^2 \left( H - \frac{T_M}{S} \right) + w_L A_L}{w_M S^2 + w_L}$$

$$(8) H = \frac{w_M S^2 \left( L + \frac{T_M}{S} \right) + w_H A_H}{w_M S^2 + w_H}$$

$$(9) S = \frac{w_M D^2 \frac{T_M}{D} + w_S T_S}{w_M D^2 + w_S}, \quad M = DS$$

The relationships for L and H (in (7) and (8)) are similar to those derived from the AD model ((3) in Section 4.1).  $T_M/S$  is the duration that would yield the target magnitude  $T_M$ . Thus, (8) means that the timing of H is the weighted average of  $A_H$  and the duration that would yield the target magnitude  $T_M$ . Only, the weighting now depends on  $S^2$ . The same holds for the timing of L in (7). As for slope, in (9),  $T_M/D$  is the slope that would yield the target magnitude  $T_M$ . Thus, (9) predicts S to be the weighted average of target slope  $T_S$  and the slope that would yield the target magnitude  $T_M$ , where the weighting depends on  $D^2$ .

The model parameters were estimated by fitting these relationships to the actual data. The relationships are non-linear (because there are variables in the denominators), so the R package *nlme* for non-linear mixed-effects models (NLME) was used for model fitting. To begin,  $w_M$  was set to 1 because only the ratios of the constraint weights are relevant, so one constraint can be fixed. We set  $T_M$  to be the average magnitude from the actual data (51Hz), because otherwise there were too many parameters for the model to converge, given the data. Thus, the model parameters obtained from the following procedure are not necessarily the optimal ones, but should be considered as an illustration of one of the parameter sets that conform to the data. There exist reasonable parameter sets for the given data, once  $T_M$  is fixed.

We started with fitting the L model in (7), to re-estimate the anchor.  $A_L$  was hypothesized to be in the first or second syllables. The L model with  $A_L$  in the second syllable was better (deviance of 3544) than the one with  $A_L$  in the first syllable (deviance of 4051). Based on this,  $A_L$  is estimated as  $A_L = v_2 + 0.12 \cdot \text{rime}_2$ . This estimate is close to the actual timing of L. The mean of the new estimate is 224, the mean of the L is 229, whereas the mean of the previous  $A_L$  was 34 (with the addition of offset (194) to this, this value becomes 228). Thus, the SAM model yielded a more reasonable estimate of  $A_L$  without the offset. Based on this model,  $w_L = 0.21$ .

The anchor for H was also re-estimated as  $A_H = v_2 + 1.03 \cdot \text{rime}_2$ . This is again closer to the actual H timing than the previous estimate. The mean of the actual H timing was 367, the mean of the previous estimate ( $A_H = v_2 + 0.93 \cdot \text{rim}_2$ ) was 333 (difference of 34), and the mean of the current estimate is 347 (difference of 20). From the H model,  $w_H = 0.15$ . Next, the NLME was fitted to the data for the S model in (9) to obtain  $w_S$  and  $T_S$ . From these fitted models, we obtained the following parameter values.

## (10) The parameters in the SAM model

$$\begin{aligned}
w_M &= 1 \\
w_L &= 0.21 \\
w_H &= 0.15 \\
w_S &= 12255 \\
T_M &= 51 \text{ Hz} \\
T_S &= 0.41 \text{ Hz/ms} \\
A_L &= v2 + 0.12 \cdot \text{rime}^2 \\
A_H &= v2 + 1.03 \cdot \text{rime}^2
\end{aligned}$$

The ratio between  $w_L$  and  $w_H$  is different from the result from the AD model. In the AD model,  $w_H$  was about three times higher than  $w_L$  ( $w_L=0.15$ ,  $w_H=0.49$ ). However, the new estimates show that  $w_L$  is higher than  $w_H$ . This means that the alignment of L is less affected by speech rate. In fact, it has been noted that in other languages, the L is more stably aligned than the H (Spanish, Prieto and Torreira 2007; Dutch, Caspers and van Heuven 1993; Seoul Korean, Cho 2011). This may be due to differences in gestural coordination between segmental and suprasegmental gestures at the beginning and the end of pitch rises (Prieto and Torreira 2007:491), but when the L is associated with the phrase-initial syllable (e.g. Seoul Korean), it may also be simply due to the fact that there is more time available for H to vary than for L because the L cannot precede the phrase onset. However, within the associated syllable, the L is shifted slightly earlier, away from the H, under time pressure (Cho 2011). Also, in Dutch, the L is shifted earlier (within the accented syllable), along with the H, when the adjacent tone is close (Caspers and van Heuven 1993:169). Thus, the timing of both L and H are subject to variation, but to different degrees, which is reflected in the constraint weights.

We compared the AIC values between the AD and the SAM models, for the goodness of fit. The AIC values were lower in the SAM model than in the AD model only for L: in the SAM model, 3558 (L), 3503 (H), in the AD model, 3722 (L), 3418 (H). An advantage of the SAM model is that it models the relationship between magnitude and slope of the rises, which the AD model cannot, but there is room for improvement in the SAM model.

Note that this is a result when  $T_M$  is fixed. The AIC of the H model from the SAM model is lower than that of the H model from the AD model if we estimate  $T_M$  to maximize goodness of fit of each model. With the addition of a  $T_M$  fixed effect and by-speaker random intercepts for  $T_M$ , the AIC of the H from the SAM model is 3385, which is lower than the one from the AD model (3418). Moreover, if we estimate  $T_M$  in fitting the L model as well, the estimates of  $T_M$  from the two models are consistent. That is,  $T_M$  from the L model was 65,  $T_M$  from the H model was 76, but they are not significantly different. The t-test shows that  $T_M$  from L and  $T_M$  from H were not significantly different ( $t(368)=0.81$ ,  $p=0.42$ ). The average of these



two estimates is  $T_M=71\text{Hz}$  (cf. the observed average magnitude is 51Hz). From  $T_M$  and  $T_S$ , the duration of the rise that satisfies both targets can be calculated: this duration is 173 ms. This is comparable with the duration target from the AD model ( $T_{DH}=176\text{ ms}$ ) (Section 4.2.3). In sum, the SAM model for the timing of L and H fit the data better than the AD models, as indicated by lower AICs and give reasonable parameter estimates. Thus, substituting the duration target  $T_D$  with  $T_M/S$  is an improvement over the AD model. However, the S model does not converge if  $T_M$  is allowed to vary or set to the  $T_M$  estimate (71) from the L and H models. This suggests that the form of the slope constraint is not correct, or that there is some additional constraint influencing slope.

One possible source of errors of the fitted SAM model with  $T_M$  varied is allowing random effects depending on speaker, when the speaker-dependent variation is large. Speaker-specific values of  $T_M$  are reasonable, because different speakers have different pitch ranges. The problem is if the speaker-specific values in the L and H models are inconsistent. In the L model,  $T_M$  varied as much as 54 Hz, and in the H model,  $T_M$  varied as much as 30 Hz. Given that the estimated  $T_M$  is 71 Hz and the actual mean is 51 Hz, these variances are quite large. Still, when  $T_M$  is not fixed, the L and H models are better in the SAM model than in the AD model: the AIC values of the SAM model with the by-speaker random intercepts for  $T_M$  were lower (3506(L), 3385(H)) than those of the AD model (3722 (L), 3418 (H)). The AIC values of the SAM model without the by-speaker random intercepts for  $T_M$  were also lower (3548(L), 3397(H)) than those of the AD model.

The weight of the slope constraint ( $w_s$ ) in (10) seems large, but this could just reflect the fact that slopes are very small numbers compared to magnitudes and durations, as a consequence of the units used. The target slope is 0.41 Hz/ms. Thus, these model parameters show that in the English L+H\* pitch accent, not only the tonal alignment, but also the slope and magnitude targets are important.

## 6. Conclusion

In this paper, we have examined the timing of tonal targets in the English L+H\* pitch accent. With varying time pressure, tendencies to maintain both segmental anchoring and target duration were observed. In Section 4.1, we developed the model using weighted constraints (Flemming 2001). The model has constraints for alignment and duration targets (the Alignment-Duration model). According to the AD model, the actual timing of L and H tonal targets are determined as the values that minimize the summed cost of violations of the alignment and duration constraints. The contribution of the duration target was significant (Section 4.4). In Section 5, we showed that the duration target in fact reflects the slope and magnitude targets of the rise.

Our findings suggest that the English L+H\* pitch accent has a target for its shape. According to the AD model, the constraint weight for the duration target ( $w_D$ ) was quite high. The relative constraint weights from the AD model indicated that satisfying the peak alignment is the most important, but satisfying the duration constraint is also important. We also applied the model with slope, alignment, and magnitude (the SAM model, Cho and Flemming 2011) to our experimental data. In the SAM model, the duration target is expressed in terms of the magnitude and slope of the rises. Model comparison indicates that the SAM model fits better than the AD model for the L and H timing. Our findings suggest that not only the alignment, but also the targets for slope and magnitude are important in the English pitch accent. To sum up, the phonetic realization of the English L+H\* pitch accent must be specified for not only alignment, but also shape targets such as slope and magnitude of the rise.

#### Appendix: Speech materials

1. Amelia Raymond
2. amenable meaning
3. eliminate Mariners
4. illuminated mirrors
5. analogous analysis
6. anonymous writings
7. anomalous meaning
8. immoral ambition
9. Norwegian marinas (for E1, E2)/ remaining minutes (for E3, E4)
10. Orwellian nightmare
11. immunity reaction
12. unmanageable employee
13. unruly employee (for E1, E2)/ Armani employees (for E3, E4)
14. linoleum knives
15. Millennium Resort
16. maligned by the media
17. remunerate a lawyer
18. malaria in Nigeria
19. aluminum mini blinds

#### REFERENCES

- ARVANITI, AMALIA and GINA GARDING. 2007. Dialectal variation in the rising accents of American English. In J.Cole and J.H.Huale (eds.). *Papers in Laboratory Phonology 9*, 547-576. Berlin, New York: Mouton de Gruyter.

- ARVANITI, AMALIA, D. ROBERT LADD, and INEKE MENNEN. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26, 3-25.
- ATTERER, MICHAELA and D. ROBERT LADD. 2004. On the phonetics and phonology of “segmental anchoring” of F0: evidence from German. *Journal of Phonetics* 32, 177-197.
- BOERSMA, PAUL and DAVID WEENINK. 2010. Praat. doing phonetics by computer [Computer program]. Version 5.1.44.
- BRUGOS, ALEJNA, STEFANIE SHATTUCK-HUFNAGEL, and NANETTE VEILLEUX. 2006. Transcribing prosodic structure of spoken utterances with ToBI. MIT Open Courseware. 6.911.
- BURNHAM, KENNETH P. and DAVID R. ANDERSON. 2002. *Model selection and multimodel inference. A practical information-theoretic approach*. New York: Springer.
- CHO, HYESUN. 2011. The timing of boundary marking tones in Seoul Korean and a weighted-constraint model. Manuscript submitted for publication.
- CHO, HYESUN and EDWARD FLEMMING. 2011. The phonetic specification of contour tones: the rising tone in Mandarin. In *Proceeding of the 17<sup>th</sup> International Congress of Phonetic Sciences*, Hong Kong.
- CASPERS, JOHANNEKE and VINCENT J. VAN HEUVEN. 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50, 161-171.
- DILLEY, LAURA C., D. ROBERT LADD, and ASTRID SCHEPMAN. 2005. Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics* 33, 115-110.
- FLEMMING, EDWARD. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18.1, 7-44.
- FOUGERON, CÉCILE and SUN-AH JUN. 1996. Rate effects on French intonation: prosodic organization and phonetic realization. *Journal of Phonetics* 26, 45-69.
- GOLDSMITH, JOHN A. 1976. *Autosegmental Phonology*. PhD Dissertation. MIT.
- GELMAN, ANDREW and JENNIFER HILL. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- GREENBERG, STEVEN and ERIC ZEE. 1997. On the perception of contour tones. Revised version of a paper presented at the 94<sup>th</sup> meeting of the Acoustical Society of America, Miami Beach, Florida.
- IGARASHI, YOSUKE. 2004. “Segmental Anchoring” of F0 Under Changes in Speech Rate: Evidence from Russian. *Proceedings of the International Conference: Speech Prosody 2004*, 25-28.
- LADD, D. ROBERT. 2004b. Segmental anchoring of pitch movements: autosegmental phonology or speech production? In Hugo Quené and Vincent van Heuven (eds.). *On Speech and Language: Essays for Sieb B. Nooteboom*, 123-131. LOT.

- LADD, D. ROBERT and ASTRID SCHEPMAN. 2003. "Sagging transitions" between high pitch accents in English experimental evidence. *Journal of Phonetics* 31, 81-112.
- LADD, D. ROBERT, DAN FAULKNER, HANNEKE FAULKNER, and ASTRID SCHEPMAN. 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of Acoustical Society of America*, 106 (3), 1543-1554.
- NIBERT, HOLLY J. 2000. *Phonetic and Phonological Evidence for Intermediate Phrasing in Spanish Intonation*. PhD Dissertation. University of Illinois.
- PIERREHUMBERT, JANET B. 1980. *The Phonology and Phonetics of English Intonation*. PhD Dissertation. MIT.
- PRIETO, PRILAR and FRANCISCO TORREIRA. 2007. The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing Spanish. *Journal of Phonetics* 35, 473-500.
- PRINCE, ALAN S. and PAUL SMOLENSKY. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. The MIT Press, Cambridge, MA. [Published version of Prince and Smolensky 1993.]
- SCHEPMAN, ASTRID, ROBIN LICKLEY, and D. ROBERT LADD. 2006. Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34, 1-28.
- SILVERMAN, KIM and JANET PIERREHUMBERT. 1990. The timing of prenuclear high accents in English. In John Kingston and Mary Beckman (eds.). *Papers in Laboratory Phonology I*, 72-106. Cambridge University Press.
- SILVERMAN, KIM, MARY BECKMAN, JOHN PITRELI, MORI OSTENDORF, COLIN WIGHTMAN, PATTI PRICE, JANET PIERREHUMBERT, and JULIA HIRSCHBERT. 1992. TOBI: A standard for labeling English Prosody. In *Proceedings of ICSLP92*, volume 2, 867-870.
- STEELE, SHIRLEY A. 1986. Nuclear accent F0 peak location: Effects of rate, vowel, and number of following syllables. *Journal of Acoustical Society of America* 80, S51-S51.
- SUNDBERG, JOHAN. 1979. Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7, 71-79.
- WELBY, PAULINE. 2006. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34, 343-371.
- THYER, NICK and DOUG MAHAR. 2006. Discrimination of nonlinear frequency glides. *Journal of Acoustical Society of America* 119, 2929-2936.
- XU, YI and XUEJING SUN. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of Acoustical Society of America*, 111 (3), 1399-1413.

Hyesun Cho  
Department of Linguistics  
Seoul National University  
1 Gwanak-ro, Gwanak-gu, Seoul 151-745, Korea  
E-mail: chohazel@gmail.com

received: July 4, 2011  
accepted: August 15, 2011