The present study examines the behavior of the lower edge of the tonal space (i.e., the L points or minima found between $F_0$ peaks) in Spanish downstepping $F_0$ contours as a function of the following four linguistic factors: (1) the $F_0$ height of surrounding peaks, (2) the temporal interval from L points to preceding peaks, (3) phrasal length, and (4) phrasal position. In contrast with H values, which are well predicted by a constant ratio of downstep from the previous H peak (Prieto, Shih & Nibert, 1996), L values are shown to be more variable and affected by a variety of contextual factors. The current study reports the results for data collected from three speakers of Mexican Spanish. An adequate fit to the data of all the speakers (between 55% and 80% of the variance explained) can be obtained by multiple regression when all these contextual factors are included. Explicit quantitative comparisons strongly suggest that such a full linear model is superior both to linear models that exclude any of these variables and to models that fit a constant-slope falling line from the preceding peak to the L value. Finally, it is found that the L values in utterance-final valleys are also affected by final lowering, suggesting that LH pairs behave as domain units for the application of phonological downstep.

1. Introduction

Throughout the history of the study of the $F_0$ implementation of intonational contours, most of the work has concentrated on the scaling of the upper part of the tonal space (or local band of $F_0$ values delimited by local peaks and valleys).¹ At least within one tradition it is widely accepted that the height of high $F_0$ peaks in stepping contours can be successfully predicted by using a constant ratio of decay from the previous peak’s value (or downstep) in languages such as English, Dutch, Swedish, Japanese, and Spanish.

¹ Ladd (1990) defines tonal space or register as a “band of $F_0$ values, a subset of the full range relative to which local targets are fixed”: The top and bottom of the register are the local default values of H and L tones, respectively, while the middle of the register is the local default value for a neutral pitch from which $F_0$ movements to H or L begin, and to which $F_0$ tends to return after H or L.
In contrast, the study of the behavior of the lower part of the tonal space (or L tone line) has been somewhat neglected in the literature, and many questions remain to be answered before we can evaluate the various models of tone scaling that have been proposed in the last two decades.

Recent analyses of peak accents in English and Japanese seem to suggest that articulatory control of H values might be different from that of trailing Ls and that we might be facing a scaling problem that departs from that of H peaks (Beckman & Pierrehumbert, 1992; Kubozono, 1993). The fact that the amount of dipping tends to increase as the temporal distance between surrounding peaks increases in English suggests that factors other than a proportional lowering from previous Hs are playing an important role in L scaling (Pierrehumbert, 1980; Anderson, Pierrehumbert & Liberman, 1984, for English; and Kubozono, 1987, 1993, for Japanese). Yet contradictory results apparently come from the study of L values of Greek prenuclear accents (Arvaniti & Ladd, 1995; Arvaniti, Ladd & Mennen, 1998): $F_0$ valleys in between peaks are found to be very stable in terms of both scaling and alignment (they occur approximately 5 ms before the onset of the accented syllable, and their height is not affected by the number of unaccented syllables intervening between accents).

The present study explores the behavior of the L tone line in Spanish downstepping contours, and represents a follow-up on a previous investigation dealing with peak height (H tone) prediction in the Spanish data collected by Prieto, Shih & Nibert (1996). In that study, peak height could be reasonably well modelled as a constant proportion of decay from the preceding peak’s $F_0$ value: between 65% and 80% of the variance of the data could be explained by using the height of the previous peak as a main predictor.³ The present study examines whether L scaling is primarily dependent on the height of the previous peak (as in H scaling) or do other factors like phrasal length or distance to surrounding peak strongly affect its values? Do we have evidence that L values behave as targets? What should be the actual domain of downstep? Does the behavior of the L line provide any evidence for downstep of the tonal space or register shift, as Ladd (1990) suggests? The goal of this investigation is to provide a preliminary descriptive model of L line scaling in Spanish and to discuss the results obtained in the light of recent proposals of tonal implementation.

2. Experimental design

The scaling of $F_0$ valleys in read declarative utterances will be examined as a function of the following linguistic variables: (1) $F_0$ height of surrounding peaks, (2) temporal

² The standard view of downstep (Liberman & Pierrehumbert, 1984; Pierrehumbert, 1980) can be represented by the following equation:

\[ P(x + 1) = r (P(x) - R) + R, \]

where $P(x)$ is the frequency of a peak in position $x$, $r$ is the fixed downstep factor and $R$ is the reference line for a particular subject. Proponents of the “superpositional models” (e.g., H. Fujisaki, E. Garding, N. Gronnum, B. Möbius, D. O’Shaughnessy) do not agree with this view of downtrend.

³ The statistical adequacy of previous peak height as a predictor of following H values will be revisited below.
distance to the preceding peak (in ms and in number of unstressed syllables), (3) sentence length (in terms of number of pitch accents per sentence), and (4) sentence position.

Fig. 1 illustrates the $F_0$ contour under study as produced by speaker JC. The utterance is a possible reading of a neutral declarative sentence and displays an overall downward trend across a series of local accent peaks. It could be described as a series of high accents or simple peaks ($H^*$), such that each peak is lower than the one preceding it in the sequence; in this context, valleys are interpreted as just “sags” between peaks. (We have avoided the transcription $L^#H^*$, with a separate $L$ target for the beginning of the rise, because we do not have evidence for a phonological contrast between $H^*$ and $L + H^*$ accents in this variety of Spanish.)

2.1. Test utterances

Thirty target utterances were constructed consisting of all possible combinations of 2 to 5 pitch-accented syllables separated by 2 to 3 intervening unstressed syllables. The following utterances (shown in Spanish orthography but also marking lexically stressed syllables content words) illustrate how the number of intervening unstressed syllables is systematically increased for 2 and 3-accent phrases (e.g., rayo de luna vs. rayo de la luna):

2 ACCENTS

rayo de luna
‘ray of moon’

rayo de la luna
‘ray of the moon’

3 ACCENTS

rayo de luna de mayo
‘ray of moon of May’

rayo de luna de mi mayo
‘ray of moon of my May’

rayo de la luna de mayo
‘ray of the moon of May’

rayo de la luna de mi mayo
‘ray of the moon of my May’

Other varieties of Latin American Spanish have been reported to have a phonological distinction between a $H^* + L$ and a $H^*$ accent: in Caracas Spanish, for example, while $H^*$ is employed in neutral declaratives, $H^* + L$ is used to convey a meaning of complaint or resentment (Sosa, 1991, pp. 15, 110).
As shown above, target utterances are complex noun phrases consisting of a head noun (Rayo de luna de mayo de gala de Lola) modified by an increasing number of following prepositional phrases (de luna, de mayo). To facilitate comparisons of pitch accents in the same position across phrases of different lengths, the ordering of content words in each phrase is kept constant and the placement of the accent is always on the first syllable of a bisyllabic noun. Finally, to minimize segmental effects on $F_0$, all content words (except the word gala) contain only sonorant consonants. The complete set of phrases is listed in the Appendix.

2.2. Recording and measurements

Three male speakers coming from two different regions of Mexico (JC and AH from Mexico City and RS from Ciudad Juárez, located in the northern state of Chihuahua) read the randomized target set of declarative utterances, for a total of 540 (30 sentences x 6 repetitions of each phrase x 3 subjects). They were instructed to read the target phrases at a normal speech rate, in a single intonational phrase, and avoiding emphasized constituents. Spoken versions of the utterance were not presented to the readers before producing the phrases.

In general, the three speakers pronounced the utterances with a continuously descending $F_0$ contour. The only difference found across subjects was the phonetic realization of the sentence-final accent: While speakers JC and AH produced very compressed final H* accents (see Fig. 1), speaker RS realized the final accent as a continuously decreasing slope starting from the peak of the penultimate pitch accent to the end of the utterance (see Fig. 2).

After recording, the experimenter listened to the utterances and discarded the ones containing discontinuities, disruptive pronunciations, or unsolicited prominence effects (clearly perceived contrastive accents on any of the constituents). For each utterance, the $F_0$ contour was produced using the Waves speech analysis package (Talkin, 1989). Relevant
F₀ measures were taken avoiding obvious tracking errors and segmental disruptive effects on the F₀ curve. Finally, the F₀ values of the labelled points were plotted to produce graphs, and any remaining F₀ tracking errors revealed in the graphs were corrected.

The labeling scheme included the following key points on the F₀ contour (see the bottom part of Figs. 1 and 2, which illustrate the marks in the label file with two five-accent phrases):

1. Phrase-initial F₀ value (marked with L)
2. The following lowest F₀ value, before F₀ starts to rise (L0)
3. Highest absolute F₀ value for each pitch accent (H₁, H₂, H₃, ...)
4. Lowest absolute F₀ value (or “valley”) between successive highs (L₁, L₂, L₃, ...)
5. Phrase-final F₀ value (L)

In the labeling scheme, H and L marks do not refer to phonological elements but to absolute locations of F₀ peaks and valleys. In general, valleys are aligned with syllable onsets and peaks are slightly “delayed” and located after the end of the accented syllable’s boundary (see Prieto, van Santen & Hirschberg, 1995). The consistently different shape of RS’s final accent triggered a redefinition of labels in this position, as this speaker produced final accents as a continuously falling slope, with no clear peak (see, for example, the absence of L₄/H₅ values for this speaker in Fig. 4). It thus made no sense to employ the H label to refer to the highest point of the F₀ peak (since no peak was present). In his case (Fig. 2), the last H corresponds to the onset of the accented syllable and the penultimate L corresponds to the end of the accented vowel. Thus, for RS’s final accent, the labels H and L are aligned to important segmental boundaries in the phrase-final accented syllable rather than to specific points in the F₀ contour. Due to their divergent behavior, RS’s final accents will be excluded from all analyses.

3. Results

3.1. Utterance-initial and utterance-final F₀ values

Starting and ending F₀ values were found to be nearly constant for a given speaker. Fig. 3 shows the mean values of utterance-initial (left panel) and utterance-final F₀ values (right panel) in phrases of increasing lengths (from 2 to 5 pitch accents). Both the initial and final values are more or less invariant for the three speakers, with the initial F₀ values displaying slightly more variation and a tendency to increase with sentence length. Despite this tendency, t-tests comparing both the values of initial and final values across phrases of different lengths were non-significant (p > 0.05) for all three speakers.

The near-constancy of utterance-final F₀ values for a given speaker has been repeatedly found in English (Maeda, 1976; Menn & Boyce, 1982; Liberman & Pierrehumbert, 1984, p. 181). In Liberman and Pierrehumbert’s study, utterance-final F₀ values were totally uncorrelated with utterance length and with the values of F₀ peaks (pronounced with different degrees of emphasis). In contrast, Thorsen’s results for Danish indicate that ending points in declarative sentences are lower in longer utterances, and that they tend to remain constant in utterances with more than 4 stress groups. Thorsen suggests that this “is probably a reflection of a physiological constraint” (Thorsen, 1980, p. 16; Thorsen, 1981, p. 42). Her idea is that each speaker has a lower limit to his/her F₀ range, which might be reached in utterances of sufficient length.
Even though the scaling of utterance-initial values has been relatively less well studied across languages, what work there is also tends to show a near-constancy effect (Liberman & Pierrehumbert 1984, for English; Prieto & Shih, 1995, for Spanish).

3.2. Low line scaling

We turn now to the study of non-terminal L tones as a function of a variety of contextual factors. The plots in Fig. 4 display the mean H and L values in different sentential positions (H1 through H5: H LINE; and L0 through L4: L LINE), for each of the three speakers, in two different conditions: 2 versus 3 unstressed syllables preceding the target L values (NUMSYLS = 2 and NUMSYLS = 3, solid vs. dotted lines, respectively).

The three plots in Fig. 4 show how the two lines connecting the upper and lower half of the tonal space (HIGH LINE and LOW LINE, respectively) have a clear decay pattern. L values have a tendency to go down as the utterance progresses, being realized no lower than the baseline (or the near-constant utterance-final F₀ value) and no higher than the topline (the value of the first peak), for the three speakers. Like the H tone line, the L tone line tends to decay more dramatically at the beginning than later in the sentence. Both lines come closer together as the sentence progresses, showing the tendency of pitch range to compress over the course of the utterance. Also, speakers JC and AH show a greater fall in utterance-final valleys and H values (e.g., L4/H5 value), which seems to indicate that final lowering does not affect the last H value exclusively but both LH values of the final pitch accent. Finally, by comparing the solid lines with the dotted lines we observe that the effect of time distance between peaks on H and L scaling is relatively stronger in the case of L points. The standard error bars show that, in contrast with H values, L values in the two conditions (NUMSYLS = 2 and NUMSYLS = 3) constitute two slightly different populations in some cases.

3.3. Effects of preceding and following H values

Existing models on the implementation of downstep in English (Ladd, 1990) and Japanese (Pierrehumbert & Beckman, 1988) share the assumption that L values behave...
Figure 4. Mean $F_0$ values of successive peaks (high line) and successive L values (low line) corresponding to an increase in the number of unstressed syllables between pitch accents (NUMSYLS = 2 and NUMSYLS = 3) for the three speakers. Solid lines link mean H and L values separated by 2 unstressed syllables and dotted lines link those separated by 3 unstressed syllables. The height of the bars represents standard error values.

as targets that are scaled relative to the value of preceding peaks. Indeed, our data display a high correlation across phrasal positions between L points in a given position in the sentence and preceding H values. But the downstepping models above would also predict correlations within a given position on a token-to-token basis. To examine this, we separated the data according to phrasal position. (Obviously, not separating the input by phrasal position would result in a high correlation between the height of the peak and the following valley, given the joint tendency of both peaks and valleys to downstep
Figure 5. L1 and L2 values in Hz (top and bottom panels, respectively) as a function of the height of the previous peak (H1 and H2, respectively). The different plotting symbols represent the data for the 3 speakers (R = RS, J = JC, and A = AH).

along the sentence.) The six graphs in Fig. 5 plot the observed L-valley values (in Hz) in the first and second positions of the sentence (L1, with plots in top row; and L2, with plots in bottom row) as a function of the F0 height of the preceding peak (H1 and H2, respectively). The main reason to select the first two valley positions for this analysis is that the size of these populations is considerably larger than those for L3 and L4. In each graph, the three subjects are represented by different plotting characters (R = RS, with plots in left column; J = JC, in center column; and A = AH, in right column). The three regression lines (represented by solid lines) summarize the high correlations observed ($R^2$ values: 0.41 for speaker RS, 0.23 for speaker JC, and 0.41 for speaker AH). In short, if a specific utterance token has a higher than average H peak in a given position in the sentence, then the following L value will also tend to be higher than average.

Similarly, the six plots in Fig. 6 also provide evidence for a strong correlation between L values and the F0 height of following H peaks ($R^2$ values: 0.46 for speaker RS, 0.45 for speaker JC, and 0.45 for speaker AH). If a L value in a given position in the sentence is higher, the prediction is that the following H peak will also increase in F0 height, and conversely.

Scaling of L points appears to be parallel to the scaling of peak values in the sense that both H and L values are highly correlated with values of preceding H points. However, the $R^2$ values for H tones in the previous study (Prieto et al., 1996) were about 9 to 19 percentage points higher for the same subjects. The question remains as to whether L values can be more accurately predicted when other factors are considered.

3.4. Effects of temporal distance to preceding peaks

The tendency for L values to lower as the time interval between surrounding peaks increases has been reported in languages like English, Japanese, and Spanish.
Pierrehumbert suggests that the degree of dipping between peaks depends on the amount of time available, as $F_0$ "falls until it is time to start aiming for the next $H^*$ level" (1980, p. 71). Kubozono (1987) also observes that valley values in Japanese, in contrast with $H$ values, are more variable and correlated with the temporal distance to preceding peaks. Finally, in a study of the effects of accentual clash in the phonetic realization of $F_0$ peaks in Spanish, Prieto & Shih (1995) find less dipping in between $H$s that are closer in time, to the point that the dip almost disappears in a tonal clash environment (two adjacent $H$s).

Table I shows the mean duration interval (in ms) between $H$-$L$ pairs (peaks and following $L$-valleys) in different positions (POS1 = $H_1$-$L_1$, POS2 = $H_2$-$L_2$, POS3 = $H_3$-$L_3$, POS4 = $H_4$-$L_4$) in phrases of different sizes (2 to 5 pitch accents), in two conditions: 2 or 3 of intervening syllables between $H$ and following $L$ ($\text{NUMSYLS} = 2$ and $\text{NUMSYLS} = 3$). As in previous analyses, RS's data is not used in the comparison between the last two accents, given its peculiar phonetic realization.

Results in Table I clearly show a significant increase in time (ranging from about 20 to 150 ms) when the number of intervening syllables changes from 2 to 3 in a given position. The relatively varied duration differences from position to position may be attributable to segmental duration effects (the segmental composition of the syllables is different according to sentential position) and to rate of speech (RS, for example, has a slower speech tempo than the other two speakers). Despite the variability just discussed, the time added by the added syllable is substantial enough to support a declination test in our data.

Reviewer D.R. Ladd has suggested that the greater variability of $L_1$ and $L_3$ compared to $L_2$ might be due to potential phrasing effects. A binary rhythm in the pronunciation of [rayo de luna][de mayo de gala][de Lola] would explain why the end of the prosodic unit [rayo de luna] ($L_2$) is longer than other word-final syllables ($L_1$ or $L_3$). However, in the data at hand I was not able to perceive binary metrical groupings and thus I am more inclined to attribute such duration differences to varied segmental composition across groups.
### Table I. Mean durational distance (in ms) between HL pairs in different positions in the sentence (POS1, POS2, POS3, POS4) with differing number of intervening unstressed syllables (NUMSYLS = 2 and NUMSYLS = 3) in phrases of different lengths (2 to 5 pitch accents)

<table>
<thead>
<tr>
<th>No. Accents</th>
<th>Interc-tone interval (ms)</th>
<th>Speaker RS</th>
<th>NUMSYLS</th>
<th>POS1</th>
<th>POS2</th>
<th>POS3</th>
<th>POS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>239.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>360.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>234.0</td>
<td>295.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>338.9</td>
<td>432.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>248.3</td>
<td>320.3</td>
<td>305.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>360.1</td>
<td>442.3</td>
<td>346.0</td>
<td></td>
</tr>
<tr>
<td>Speaker JC</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>164.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>231.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>198.4</td>
<td>207.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>249.1</td>
<td>340.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>168.1</td>
<td>212.5</td>
<td>160.5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>228.3</td>
<td>289.3</td>
<td>227.3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>192.5</td>
<td>222.4</td>
<td>173.8</td>
<td>188.1</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>238.3</td>
<td>307.5</td>
<td>247.8</td>
<td>250.7</td>
</tr>
<tr>
<td>Speaker AH</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>158.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>241.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>197.2</td>
<td>188.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>199.5</td>
<td>246.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>168.4</td>
<td>191.7</td>
<td>190.8</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>253.2</td>
<td>236.5</td>
<td>214.7</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>NUMSYLS = 2</td>
<td>192.5</td>
<td>195.2</td>
<td>202.6</td>
<td>186.7</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>NUMSYLS = 3</td>
<td>241.7</td>
<td>247.4</td>
<td>228.0</td>
<td>233.5</td>
</tr>
</tbody>
</table>

For the declination test, the data were grouped according to phrasal position (L1, L2, L3, L4), distance in syllables between two surrounding peaks (NUMSYLS = 2 and NUMSYLS = 3), and utterance length (2 to 5 pitch accents). The effect of the number of unstressed syllables on L values is consistent in the contours produced by our three speakers: In the same phrasal position, the amount of dipping between 2 peaks increases as the number of syllables between those peaks increases (from 2 to 3). For the 3 speakers, t-tests comparing low values in the same utterance position (separated by 2 vs. 3 intervening syllables) resulted in highly significant differences in 8 out of 11 possible cases (at p < 0.05). The only non-significant comparisons were L2 for speaker RS, and L2 and L4 for speaker AH. Similarly, we performed a t-test analysis comparing all of the low values in our data in the two- and three-syllable conditions. For our three subjects, the two populations were highly distinct (at p < 0.001).  

---

7 Results of t-tests performed on all L values with 2 vs. 3-syllable conditions, for the three subjects (RS, JC and AH in that order) were t(1070) = 9.00, t(538) = 7.62, t(586) = 3.84. All tests were highly significant (p < 0.001).
As expected, the data also display a high correlation between the $F_0$ interval between adjacent HL pairs and the distance in time (in ms) between both points. The three plots in Fig. 7 show the $F_0$ interval (in Hz) observed between H1 and L1 as a function of the time interval between the two. Different plotting symbols (‘1’ and ‘3’) identify the data points separated by 2 and 3 unstressed syllables, respectively. Clearly, points separated by 2 and 3 unstressed syllables are located in separate regions along the time axis (x-axis). The regression lines in each plot summarize the positive correlation between $F_0$ and time distance: In all phrasal positions, the interval in Hz between a peak and the next valley increases as the distance in absolute time between the two increases. Still, the substantial scattering of the data points around the regression line may indicate that other variables should be considered in the prediction of L scaling.

In sum, both the syllable-based and time-based analyses performed in this section demonstrate that time distance (either in milliseconds or in number of unstressed syllables) has a significant effect on L scaling. Valleys in a given position in the sentence tend to progressively decrease as the distance in time to the preceding H peak increases.

### 3.5. Effects of phrasal length

To our knowledge, no study has reported significant effects of phrasal length on the scaling of the lower part of the tonal space.\(^8\) The two plots in Fig. 8 show the absolute

---

8 Grennum (1986) has presented data about the effects of phrasal length on declination. In her data, L values are sensitive not only to the number of sentences that make up the text but also to the length of individual sentence components. Still, we leave her data out of detailed consideration because she deals with L* or L* + H accents, whereas we are dealing with H* (or L + H*).
mean $F_0$ value (in Hz) of L points in different phrasal positions (L1 through L4) in utterances of different lengths (2 to 5 pitch accents, represented by different line types in the plot), pooling across the time distance factor. At first glance, phrase-medial L values (L2 in 4- and 5-accent sentences and L3 in 5-accent sentences) seem not to be appreciably affected by phrasal position: L valleys in this mid-sentence position attain about the same $F_0$ level, regardless of sentence length. The plots of speakers JC and AH (the two who produced utterance-final accents as compressed peaks) reveal a very interesting behavior of L values corresponding to the last pitch accent in the utterance. Consistently, when L values in the same position relative to the beginning of the utterance are compared, those that are before the final peak H are lower than those that do not precede final accent peaks. For example, the final valley of a 4-accent sentence (L3 linked by dashed lines in plot) is significantly lower than L3 in non-final positions (e.g., in a 5-accent sentence).

We interpret this phenomenon as a phonetic manifestation of final lowering, another source of downtrend affecting the last portion of an intonational phrase. This effect has been repeatedly reported of intonational contours of Japanese (Pierrehumbert & Beckman, 1988), Danish (Thorsen, 1981), German (Möbius, 1993, p. 129), English (Liberman & Pierrehumbert, 1984, p. 186), and Spanish (Prieto et al., 1996). These studies have found that the scope of final lowering varies depending on the language; the domains reported range from the last half second in English to the last H value for Spanish (both for declarative sentences). In a previous investigation (Prieto et al., 1996), we found evidence of final lowering on sentence-final H peaks, which could be accurately predicted through a separate downstep/lowering factor. What Fig. 8 shows is that L values preceding those peaks are also affected by a final lowering
phenomenon, and such behavior seems to call for a unified treatment of pairs of LH values.\(^9\)

4. A linear model of L height

Results from preceding sections indicate that the F\(_0\) level attained by L values is significantly affected by the contextual factors investigated: (1) phrasal length (NUMACC); (2) temporal interval to the preceding peak (NUMSYLS); (3) phrasal position (POSENT); and (4) height of the preceding and following peak (PREVPEAK and FOLPEAK). We will attempt to sort out the contribution of these variables by means of regression analysis. The equations below provided a good fit to the non-final data for each of the subjects. (Because final L values are affected by final lowering, those points have been excluded from this analysis. By treating the last pitch accent separately, the performance is significantly increased.\(^{10}\)

\[
\text{Low (RS)} = 44.863 - 0.085 \text{NUMACC} - 0.334 \text{POSENT} \\
- 3.585 \text{NUMSYLS} + 0.259 \text{PREVPEAK} + 0.279 \text{FOLPEAK},
\]

\[
\text{Low (JC)} = 68.802 - 1.334 \text{NUMACC} - 2.032 \text{POSENT} \\
- 5.765 \text{NUMSYLS} + 0.199 \text{PREVPEAK} + 0.309 \text{FOLPEAK},
\]

\[
\text{Low (AH)} = 39.431 - 0.475 \text{NUMACC} - 0.201 \text{POSENT} \\
- 1.555 \text{NUMSYLS} + 0.237 \text{PREVPEAK} + 0.25 \text{FOLPEAK},
\]

where NUMACC is the number of pitch accents in the sentence (2 to 5), POSSENT is the position of the accent in the sentence (2 to 5), NUMSYLS is the number of unstressed syllables preceding the accent, PREVPEAK is the F\(_0\) height of the previous peak, and FOLPEAK is the F\(_0\) height of the following peak.

The regression analysis above accounts for 81.5\% of the variance of the data for speaker RS, 74.1\% for speaker JC, and 54.3\% for speaker AH. The values of the reduction in the percentage of variance accounted for (\(R^2\)), caused by dropping each of the variables from the full equation for each of the three speakers, are given in Table II. Not surprisingly, eliminating FOLPEAK or PREVPEAK from the equation triggers a highly significant (\(p < 0.001\)) decrease in performance ranging from 2.9\% to 5.0\% for FOLPEAK and from 1.7\% to 4.5\% for PREVPEAK. Similarly, leaving out NUMSYLS triggers a significantly worse (by 4.1\% to 9.5\%) fit for all subjects. Eliminating the positional variable POSSENT triggers a substantially smaller reduction (from less than 0.1\% to 1.1\%) and is only significant (\(p < 0.01\)) for one of the three subjects (JC). Finally, removing the phrasal length factor, NUMACC, also has a weak effect on the performance of the model, decreasing the degree of fit by from less than 0.1\% to 0.8\%. Although this difference is highly

\(^9\)Reviewer M. Beckman suggests that this fact, together with the relative timing facts, might be indicating that the phonetically more transparent analysis of these accents is L + H*. Yet, the fact that L values progressively decrease as the time between surrounding Hs increase is not a confirmation of the status of L as a real target. Because of this and the lack of evidence for a phonological contrast between H* and L + H* in this variety, we leave this question open.

\(^{10}\)The performance of the model decreased significantly when including the final L values. For speaker JC, for example, the squared correlation coefficient (\(R^2\)) decreased from 74.1\% to 57.8\%. 
TABLE II. Decrease in percent variance accounted for ($R^2$) by eliminating each of the variables for the non-final L-valley tones. Significance levels determined by associated two-tailed t-tests (df = 378 for RS, 367 for JC, and 402 for AH)

<table>
<thead>
<tr>
<th>Variable removed</th>
<th>Speaker</th>
<th>RS</th>
<th>JC</th>
<th>AH</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOLPEAK</td>
<td></td>
<td>2.9b</td>
<td>5.0b</td>
<td>3.6b</td>
</tr>
<tr>
<td>PREVPEAK</td>
<td></td>
<td>3.0b</td>
<td>1.7b</td>
<td>4.5b</td>
</tr>
<tr>
<td>NUMSYLS</td>
<td></td>
<td>7.0b</td>
<td>9.5b</td>
<td>4.1b</td>
</tr>
<tr>
<td>POSSENT</td>
<td>&lt;0.1</td>
<td></td>
<td>1.1b</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>NUMACC</td>
<td>&lt;0.1</td>
<td>0.6a</td>
<td></td>
<td>0.4</td>
</tr>
</tbody>
</table>

*p < 0.01.

**p < 0.001.

significant only for JC, it is also marginally significant for AH [$t(402) = −1.92$, $p = 0.055$].

The evidence for the relevance of PREVPEAK, FOLPEAK, and NUMSYLS is very strong. Although the effects of POSSENT and NUMACC are more variable across subjects, there is some reason to retain them pending further study. First, the coefficients for all the variables, including POSSENT and NUMACC, show consistent positive or negative signs across the three subjects. Second, for two of the three subjects (AH and JC), simultaneously dropping both POSSENT and NUMACC from the larger equation results in a highly significant (though modest in magnitude) decrease in $R^2$: a decrease of 0.8% [$F(2,402) = 7.31$, $p = 0.001$] for AH, and a decrease of 2.5% [$F(2,367) = 36.13$, $p < 0.001$] for JC. While the stability of these factors across speakers remains somewhat doubtful, we will include them in the discussion below because it is clear they are having reliable effects for at least some subjects.

Unlike H values in previous studies, L values cannot be successfully predicted as proportions of the height of the previous peak. Using only PREVPEAK and FOLPEAK as predictors leads to substantial decreases in $R^2$: by 7.0% for RS, by 12.5% for JC, and by 5.0% for AH. If PREVPEAK or FOLPEAK are used singly as predictors, reductions are greater still. For PREVPEAK and FOLPEAK, respectively, the reductions are: RS: 11.5% and 10.9%, JC: 19.1% and 17.6%, and AH: 9.3% and 10.4%. Thus, a scaling hypothesis that claims that target L points can be calculated by exclusively referring to the previous H value cannot be maintained for Spanish. Factors such as temporal distance to the preceding peak and F$_0$ values of the preceding and following peaks also have a significant role in the prediction of the L values.

4.1. Final lowering: modeling final L-valley values

To provide a descriptive model of final L-valley values and to sort out the contribution of the different variables to their prediction, we performed a multiple regression analysis. Again, RS's final accents are excluded from the analysis, as they were produced as a gradually falling slope. The linear models with the regression coefficients that give
a better fit to the data for speakers JC and AH are shown below:

Final Low (JC) = 42.853 + 0.134 prevpeak + 0.576 folpeak
− 3.87 numsylys − 2.04 numacc,

Final Low (AH) = 39.565 + 0.262 prevpeak + 0.229 folpeak
− 1.737 numsylys − 0.468 numacc,

where numacc is the number of pitch accents in the sentence (2 to 5); numsylys is the number of unstressed syllables preceding the accent; prevpeak is the F₀ height of the previous peak; and folpeak is the F₀ height of following peak.

The regression analysis above accounts for 79% and 44.8% of the variance of the data for speakers JC and AH, respectively. As shown in Table III, the regression coefficients of all factors included in the analysis were statistically significant for both subjects.

As expected, very high correlations are found between utterance-final valley values and the F₀ height of preceding peaks (mean correlation coefficients are 0.69 for speaker JC, and 0.56 for speaker AH) and following peaks (0.79 for JC, and 0.50 for AH). By only employing prevpeak as main predictor, we obtain a 30.3% drop in the R² account of the data for JC, and a 13.3% drop for AH. Using only folpeak, the amount of variance explained decreases by 16.5% for JC, and 19.7% for AH. Thus, final L values are shown to behave in a similar fashion to non-final values, in the sense that they are only consistently well-predicted when all of the linguistic variables under consideration are included.

4.2. Scaling of H-tones (revisited)

The complexity of the model for the L-tones suggests that the simple model for the scaling of H tones favored by Prieto et al. (1996) should be reconsidered. The editors point out that if subsequent H values can be predicted as a linear function of a preceding H value, then those H values can be eliminated in a linear model for the expected value of an intervening L. That is, a following H tone cannot have a linear effect

<table>
<thead>
<tr>
<th>Variable removed</th>
<th>Speaker</th>
<th>JC</th>
<th>AH</th>
</tr>
</thead>
<tbody>
<tr>
<td>folpeak</td>
<td></td>
<td>11.7&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.3&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>prevpeak</td>
<td></td>
<td>1.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>7.7&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>numsyls</td>
<td></td>
<td>4.9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8.4&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>numacc</td>
<td></td>
<td>3.5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.6&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

*<sup>a</sup>p < 0.05.
*<sup>b</sup>p < 0.01.
*<sup>c</sup>p < 0.00.
on a preceding L tone unless there is some additional factor affecting that H tone beyond the $F_0$ of the preceding H tone.

Prieto et al. (1996) found that H values were well modeled using a downstep decay pattern (55% to 65% of the variance of the data was accounted for, depending on the speaker). In fact, this simple model was more successful than a linear model including all of the other independent factors examined (utterance length, time distance between accents, syllable duration, etc.), but excluding the preceding H-value.

However, what happens if the analog of the larger linear model (including the preceding H value as well as the other factors) given at the beginning of this section is used to predict H-tone scaling? Such an analysis was examined by switching the roles of the low tone and the following H peak in the prediction equation (so that HOLPEAK is the y-variable and Low is a predictor). For all three subjects, there was in fact a significant increase in $R^2$. However, quantitatively, the improvement in prediction is substantially less for predicting H tones than for predicting L tones. For predicting L tones, the improvement in $R^2$ that results by adding all the other factors to a model using PREVPEAK alone ranges from about 9 to 19 percentage points. For predicting HOLPEAK, the improvements are only about half as large (4 to 8 percentage points).

A detailed consideration of more complex but relatively subtle scaling of H tones is beyond the scope of the present investigation. Instead, an alternative account of L-tone scaling will be examined.

### 4.3. Fixed-slope model

The goal of this section is to test the performance of another possible way to account for L values, which we will call the fixed-slope model, that is, the hypothesis that L values are just the result of a linear time interpolation probably caused by a declination effect. The three plots in Fig. 9 represent the linear slopes linking an H value in a given position in the sentence (H1 to H4) to the following L value (L1 to L4), separating the data according to differing number of intervening unstressed syllables and collapsing phrases of different lengths (from 3 to 5 pitch accents). The slopes represented in Fig. 9 link an H value to the next L value over the actual mean time distance in ms (which is represented on the x-axis). Within the same phrasal position, lines linking HL values tend to have similar falling slopes across the different conditions (see NUMSYLS = 2 and NUMSYLS = 3 conditions in the plot). If anything, slopes in the 2-syllable condition are steeper than those in the 3-syllable condition. Also, the three plots show that the slopes are slightly less steep over the course of the utterance (L1 > L2 > L3).

The fact that L values are progressively lower as the temporal interval between surrounding peaks increases provides preliminary evidence to believe that the velocity of the slope falling from the peak to the following $F_0$ valley might be similar in different positions, and thus L scaling might be better characterized as a fixed-slope line linking the preceding peak with the L value. Consider, then, the hypothesis that speakers produce falling slopes after H targets as a result of a constant declination, with no aiming at real targets (as H values). Within this approach, L points are obtained by fitting a constant slope line from the previous peak to the position of the L value, generally anchored at the onset of the accented syllable (Prieto, van Santen & Hirschberg, 1995), as indicated in the equation below. Slope constants (s) for the three subjects are obtained by dividing the mean distance in Hz between Hs and following valleys by the time distance between the two. The mean slope constants are $-0.0833$ for speaker RS, $-0.1955$ for...
Figure 9. Mean slopes connecting H-L pairs in different positions in the utterance (H1 through H4, and L1 through L4), varying the number of unstressed syllables between both NUMSYLS = 2 and NUMSYLS = 3). X-axis plots time in ms. Plotting symbols represent H and L values in different positions: ‘‘.’’ links H1 and L1, ‘‘.’’ links H2 and L2, ‘‘.’’ links H3 with L3, and ‘‘.’’ links H4 with L4.

The scaling of the L tone line

Speaker JC, and $-0.1012$ for speaker AH.

$$F_0 (L) = F_0 (\text{PrevPeak}) - s (t(\text{PrevPeak}) - t(L)), $$

where $F_0 (L)$ is the $F_0$ value of a given L, $F_0 (\text{PrevPeak})$ is the $F_0$ value of the peak preceding L, s is the slope constant, t(PrevPeak) is the time value of the peak preceding L, and t(L) is the time value of L.
To compare the performance of the fixed-slope model and the linear model of the preceding section, we computed the predictions of both models and compared the resulting $R^2$ values. The $R^2$ values for the fixed-slope model were 68% for speaker RS, 61% for speaker JC, and 44% for speaker AH. In contrast, $R^2$ for the full linear model (i.e., the one including all of the factors) is 81.5% for speaker RS, 74.1% for speaker JC, and 54.3% for speaker AH. Fig. 10 plots L values, as predicted by the fixed-slope model, as a function of the observed values, for each speaker (RS, plot on the left; JC, plot in the center; and AH, plot on the right). Solid lines represent $x = y$, and dotted lines are least square regression lines.

Note that the constant slope model tends to consistently overshoot L values in lower $F_0$ ranges (i.e., those that are further into the sentence) and to undershoot them in higher $F_0$ ranges (see the position of the actual regression line with respect to the $x = y$ line). This is attributable to the fact that slope values have the tendency to decrease as the sentence progresses (see slope plots in Fig. 9). The inaccurate predictions of the fixed-slope model seem to argue against the idea that L values are simply a result of a passive, tonally unmediated decline before beginning the approach to the next articulatory H target.\textsuperscript{11}

\textsuperscript{11}Formal statistical tests between models of different families are not readily available and the editors note that some caution must be exercised because of the larger number of parameters fitted in the case of the full model. However, non-statistical evidence is quite compelling. First, as reported in section 4.2, the full linear model shows very substantial increases (from 10% to 13%) in the variance accounted for all three subjects, despite the relatively modest number of parameters estimated. Second, the fixed-slope model shows systematic error (not present in the linear model), in that it consistently overshoots in lower $F_0$ ranges (L values that are further into the sentence) and undershoots in higher $F_0$ ranges. Also, the significance of the regression coefficients of all factors included in the linear model (NUMACC, POSSENT, NUMSYLS, PREVPEAK, FOLPEAK), as well as their positive correlation with L values (see sections 3.1 to 3.5), suggest very strongly that factors other than frequency of the previous H peak are involved in the model.
5. Conclusion

Examination of the lower part of the tonal space in Mexican Spanish downstepping contours has revealed that values for valleys between peaks are not as constant as the peak values themselves. L points, as opposed to H points, cannot be successfully predicted as a fixed proportion of preceding Hs (see Prieto et al., 1996, for a description of the high tone line using the same data). Ls are not scaled lower than the preceding or following peak by some fixed proportion or interval, but are greatly affected by a variety of linguistic factors. To begin with, L values are clearly time-dependent: Correlation analyses show that the amount of dipping in between peaks increases as the distance in time to the preceding peak increases, for the three subjects. This result confirms similar observations in other languages (Pierrehumbert, 1980; Anderson et al., 1984, for English; and Kubozono, 1987, 1993, for Japanese) and recent descriptions of Spanish tonal clash contexts, where the dip in between adjacent peaks almost disappears when it has not enough time to be realized (Prieto & Shih, 1995). The only apparently contradictory result to this observation is the work on Greek by Arvaniti & Ladd (1995) and by Arvaniti, Ladd & Mennen (1998). Both studies confirm that valleys of prenuclear accents behave as targets in both the alignment and the scaling sense. The two studies find no correlation whatsoever between the F0 interval between the H and the following L and the duration of that interval. The different L behavior between Greek and Spanish might be due to the different phonological makeup of the accents under investigation (L + H* in Greek vs. H* in Spanish).

Correlation analyses of our data also show that valleys are dependent on the height of both the preceding and following peaks. In fact, as we have already mentioned, the linear model including all of the contextual factors under study (F0 of preceding and following H peaks, distance in time to preceding H, phrasal position, and phrasal length) leads to a good fit to the data (between 55% and 80% of the variance of the data is explained). Thus, the Spanish data studied in this article proves that, even though the most significant predictor of L values is the F0 height of the previous peak, these values are predicted more successfully if we include other contextual factors in the analysis.12 Two recent models of tonal implementation proposed for two different languages (Ladd, 1990, for English; Pierrehumbert & Beckman, 1988, for Japanese) suggest that L values between Hs are scaled proportionally to high tones. Thus, both models describe L values as target points scaled in relation to preceding H peaks.13 Yet the present results show that that (at least in some languages) the effects of temporal distance to preceding peaks suggest that factors other than a proportional lowering from previous Hs are playing a crucial role in the prediction of L values. Thus, as Beckman & Pierrehumbert (1992) and Kubozono (1993) suggest, we might be facing a scaling problem that is substantially different from that of H peaks.

The fact that L points are greatly affected by the distance in time to the preceding peak leads to the hypothesis that the scaling of L values could be governed by a

---

12 As reviewers M. Beckman and D.R. Ladd point out, the two theories mentioned (Ladd, 1990; Pierrehumbert & Beckman, 1988) were designed to account for two different language systems. Moreover, both models are just in an incipient stage and more data should be examined before defining a universal model of pitch scaling. Finally, we have to keep in mind that the scaling situation described for leading L tones here might be different from that of L* accents across languages.

13 In Beckman and Pierrehumbert’s words, “It is clear that the value of the L% tone is linearly related to the value of that preceding H” (1988, p. 30).
time-dependent function, by fitting a line with a near-constant slope linking each peak with the start of the rise (generally located at the onset of the accented syllable). An explicit statistical comparison between what we called the fixed-slope model and the linear model reveals that the linear model provides a better fit to the data: The fixed-slope model consistently undershoots targets in utterance-initial positions and overshoots targets in utterance-final position. The inaccuracy of the constant-slope model seems to prove that we are not dealing with a simple linear declination effect or gradual drop in $F_0$ in between highly controlled H targets.

Thus, while we can view H peaks as highly stable and precise targets, L values are more varied and are probably displaying different degrees of undershoot of L values in between closely spaced H targets. Thus, as Beckman suggests (see note 5), the situation just described (highly targeted Hs versus undershot Ls) might reflect a difference in articulatory production between Hs and Ls that is comparable to the articulatory production of CVC sequences. Lindblom’s spectrographic study of such sequences revealed that while vowel formant targets underwent undershoot, the loci of consonants were strictly maintained. Further study of these contours under different time pressure conditions is called for.

Finally, consider how the behavior of the final part of the L tone line in Spanish bears on the domain of final lowering and downstep. In these contours, final lowering affects not only the utterance-final peak but the entire LH pair; that is, both L and H values in phrase-final position are substantially lower than non-final Ls and Hs, respectively, in the same position relative to the beginning of the utterance (see Prieto et al., 1996, for modeling of the final lowering of the peak). This type of behavior seems to support a phonological unit consisting of LH points, as both ends of the tonal space are affected by the same process of final lowering, and not just the H tonal line. The two scaling models mentioned above (Ladd, 1990, for English and Pierrehumbert & Beckman, 1988, for Japanese) disagree on what should be the domain of the downstep rule. While in Ladd’s (1990) model the downstep rule affects the whole register or both members of a LH or HL pair, Pierrehumbert & Beckman’s (1988) model predicts no dependency between LH pairs (as L values are obtained by scaling them with preceding H values). Thus, in the Spanish case at hand, Ladd’s concepts of register domain and register shift seem to capture the joint behavior of LH pairs, as they are treated as a unit by final lowering.

Part of this work was presented at the 1997 ESCA Workshop on Intonation held in Athens. I am grateful to Cinzia Avesani, Holly Nibert, Bernd Möbius, and Daniel Recasens for helpful observations, and to Julia Hirschberg, Chilin Shih, and Jan van Santen for extensive advice and discussions on the topic. (I also thank Chilin for her last-minute help on retrieving some data analyses.) I am also indebted to Mary Beckman, D.R. Ladd, an anonymous reviewer, and the editors of this journal (Terrance M. Nearay and Bruce L. Derwing) for very helpful comments and criticisms which have lead to a substantial improvement of the first draft of this paper. Finally, I would also like to thank our Mexican speakers JC, AH, and RS for their collaboration and patience during the recording sessions. This research was supported in part by an SGR grant.
no. 1997SGR00290, Generalitat de Catalunya, and DGICYT grant no. PB-96-1199-C04-04, Ministerio de Educación y Ciencia.

References


Appendix

The target utterances used in the production experiment are given in full below. Although not marked in Spanish orthography, for clarity, primary-stressed syllables are here indicated with an acute accent.

2 ACCENTS
1. rayo de luna
2. rayo de la luna

3 ACCENTS
3. rayo de luna de mayo
4. rayo de luna de mi mayo
5. rayo de mi luna de mayo
6. rayo de mi luna de mi mayo

4 ACCENTS
7. rayo de luna de mayo de gala
8. rayo de luna de mayo de la gala
9. rayo de luna de mi mayo de gala
10. rayo de luna de mi mayo de la gala
11. rayo de la luna de mayo de gala
12. rayo de la luna de mayo de la gala
13. rayo de la luna de mi mayo de gala
14. rayo de la luna de mi mayo de la gala

5 ACCENTS
15. rayo de luna de mayo de gala de Lola
16. rayo de luna de mayo de gala de la Lola
17. rayo de luna de mayo de la gala de Lola
18. rayo de luna de mayo de la gala de la Lola
19. rayo de luna de mi mayo de gala de Lola
20. rayo de luna de mi mayo de gala de la Lola
21. rayo de luna de mi mayo de la gala de Lola
22. rayo de luna de mi mayo de la gala de la Lola
23. rayo de la luna de mayo de gala de Lola
24. rayo de la luna de mayo de gala de la Lola
25. rayo de la luna de mayo de la gala de Lola
26. rayo de la luna de mayo de la gala de la Lola
27. rayo de la luna de mi mayo de gala de Lola
28. rayo de la luna de mi mayo de gala de la Lola
29. rayo de la luna de mi mayo de la gala de Lola
30. rayo de la luna de mi mayo de la gala de la Lola