This is an offprint from:

Karen Zagona (ed.)
Grammatical Theory and Romance Languages
John Benjamins Publishing Company
Amsterdam/Philadelphia
1996
(Published as Vol. 133 of the series
CURRENT ISSUES IN LINGUISTIC THEORY,
ISSN 0304-0763)

ISBN 90 272 3637 2 (Hb; Eur.) / 1-55619-588-5 (Hb; US)

THE ABSENCE OR PRESENCE OF A DECLINATION EFFECT ON THE DESCENT OF F0 PEAKS?
EVIDENCE FROM MEXICAN SPANISH

PILAR PRIETO
AT&T Bell Laboratories

HOLLY NIBERT
University of Illinois at Urbana-Champaign

CHILIN SHIH
AT&T Bell Laboratories

0. *Introduction*

The term 'declination' refers to the observation that fundamental frequency (F0) tends to decrease over the course of an utterance. It is viewed as a gradual compression of the phonetic frame of reference in which phonological F0 excursions are realized; in other words, it is a global backdrop change to local F0 events and it occurs as a function of time. Models of intonation such as the one proposed in Pierrehumbert (1980) incorporate declination into the phonetic component of a two-component mechanism. The other component is a phonological one that encompasses meaningful, local F0 movements resulting from pitch accent commands, such as the H* or L* pitch accent.

Recent studies on the scaling of descending F0 peaks (generated by a sequence of H* pitch accents) point to two different outcomes concerning the need for a time-dependent declination effect to account for this F0 contour pattern: 1) Liberman & Pierrehumbert's (1984) study on descending F0 peaks in English revealed a degree of descent too abrupt to be attributed to a time-dependent compression of the topline of the F0 range. Instead, they showed that such peaks can be successfully modelled by using a fixed ratio of decay only. Such a decay produces a local, accent-by-accent reduction of the topline of the pitch range. They found no additional descent in F0 attributable to time beyond the descent generated by a fixed ratio of decay. This work resulted in the adoption of the concept 'downstep', from work on African tone languages, to refer to the topline compression of F0 generated by fixed ratio and triggered by phonological rule in pitch accent languages as well; and 2) Since then, studies such as Pierrehumbert & Beckman (1988) on Japanese defend that peak height is best predicted by applying ratio-driven downstep on each peak and, additionally, a time-dependent declination function, once again incorporating both mechanisms.

The main goal of the present study is to seek evidence for a time-dependent declination effect on descending F0 peaks in Mexican Spanish in order to determine if this is a necessary component for their modelling in speech synthesis. To this end, we compare phonologically equivalent events across utterances—specifically, H* pitch accents to which a downstep rule has applied so that each accent's F0 peak is lower than the one preceding it—and examine if and how time further affects their F0 realization. In our data, we expect successive F0 peaks to decrease in height due to downstep; however, if declination is present, the reduction in F0 will become greater as the time interval between peaks increases. We create this time interval in terms of unstressed syllables between adjacent peaks and vary it to create three conditions: a one-, two-, or three-syllable lapse.

Our results show that the F0 descent in our contours of downstepped H* peaks is not attributable to a time-dependent declination effect. Rather, the height of a given F0 peak can be reliably predicted relative to the value of the peak preceding it by reducing this value by a fixed proportion (also called a 'downstep ratio'). These results are consistent with the findings of Liberman & Pierrehumbert (1984) for English and show that in Spanish as well, this type of descending F0 contour can be captured following a local downstep treatment. These results will be presented in detail after a brief description of the experimental design and the basic measurements used is given.

1. *Experimental design and measurements*

The database used in our experiment consisted of 120 declarative 'listing' phrases of different lengths, all comprised of a noun phrase modified by an increasing number of embedded noun or prepositional phrases. They were obtained by exhaustively combining phrases of two to five pitch accents with one to three intervening unstressed syllables. For example, the utterances shown in (1) (written orthographically, marking the lexical stress of each content word) correspond to the two- and three-accent groups. As noted

(1)

2 ACCENTS
rávo lúna
"ray moon"
rávo de lúna
"ray of moon"
rávo de la lúna
"ray of the moon"

3 ACCENTS
rávo lúna mávo
"ray moon May"
rávo lúna de mávo
"ray moon of May"
rávo lúna de mi mávo
"ray moon of my May"
rávo de lúna mávo
ray of moon May"
rávo de lúna de mávo
"ray of moon of May"

3 ACCENTS (CONT'D)
rávo de lúna de mi mávo
"ray of moon of my May"
rávo de la lúna mávo
"ray of the moon May"
rávo de la lúna de mávo
"ray of the moon of May"
rávo de la lúna de mi mávo
"ray of the moon of my May"

earlier, the number of intervening unstressed syllables was systematically increased between pitch accents, meeting all possible combinations.

The structural characteristics of the database make possible a strict analysis of the effect of the distance between pitch accents on peak height. To facilitate the comparison of peaks in the same position across the utterances, the ordering and stress pattern of content words in each phrase are kept constant. Also, all phrases are comprised of complex NPs except in the case of only one unstressed syllable between pitch accents, which results in adjacent simple NPs (e.g., *rayo luna* "ray moon"). Finally, to minimize segmental effects on peak height, all content words contain sonorant consonants only (with the exception of *gala* "gala").

A male speaker of Mexican Spanish[1] read the 120 phrases at least five times in random order, for a total of 809 utterances. The speaker was trained to read the phrases at a normal constant speech rate using a descending pitch pattern with no phrasing or emphasized constituents. The type of sentence structure (a complex noun phrase, modified by an increasing number of embedded NPs or PPs) was initially chosen because it resembles a list of elements and naturally elicits the type of downstepped contour under study. In accordance with Pierrehumbert (1980), such a contour can be characterized as a series of simple peaks, or H* accents, such that each peak is lower than the one preceding it (or 'downstepped'). The upper part of Figure 1 illustrates a typical downstepped F0 contour of an utterance with five pitch accents.

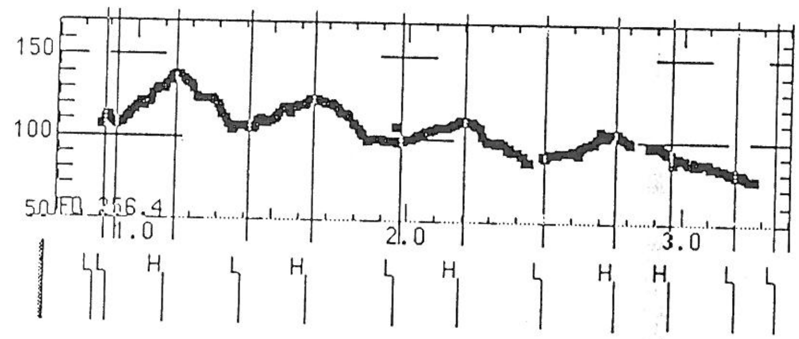

Figure 1: F0 contour of the five-accent utterance *Rayo de luna de mayo de gala de Lola*, shown here to illustrate the scheme used to label our data.

[1] RS, our speaker, is a native of Ciudad Juárez, located in the northwestern state of Chihuahua, Mexico.

After completing the recordings, each speech file was carefully checked, and utterances containing discontinuities or unsolicited prominence effects were discarded. The F0 contour of each file was extracted using the Waves speech analysis package (Talkin 1989), and relevant F0 measures were taken by manually placing tick marks or labels (avoiding F0 tracking errors and local segmental effects).

The points labeled H are those of concern in the present paper. An H marks the highest absolute F0 value of a peak. Figure 1 shows how the F0 contour of a five-accent utterance contains only four clear peaks. In the contour type elicited for this study, final pitch accents, unlike non-final ones, do not generate a clear F0 peak in the speech of our informant. Due to this divergent behavior, we excluded utterance-final pitch accents from our analyses.

2.0 *F0 peak height*

2.1 *Initial rise as evidence of constant pitch range*

In order for the F0 peaks to be comparable, they need to correspond to the same pitch accent and have been realized within a similar pitch range, since pitch range variation represents an additional factor that could confound the effect of the factors under study. First, as we can see in Table 1, initial (starting) and final (ending) F0 values (the first and last Ls in our labelling scheme) are nearly constant over utterances of different lengths: around 105 and 75 Hz, respectively.

Table 1: Mean values (in Hz) of the utterance-initial and utterance-final F0 values of phrases of different lengths (2 to 5 pitch accents).

Number of accents	Initial	SD	Final	SD
2	105.10	7.19	76.30	2.85
3	105.88	6.13	76.87	2.78
4	106.41	5.45	75.60	3.16
5	107.16	5.76	74.71	3.66

Table 2 shows that the values of the initial F0 rise (measured as the distance in Hz from the lowest F0 point at the beginning of the contour to the highest F0 point of the first peak) are more or less constant in phrases of different lengths (from three to five pitch accents). Since we know that the initial F0 values of our utterances are likewise rather stable, we can be confident that the utterances were produced over a more or less constant pitch range and that the peak values are therefore comparable.

Table 2: Mean values (in Hz) of the pitch range of the first peak in utterances of different lengths (3 to 5 pitch accents).

Number of accents	Pitch Range of Accent 1	SD
3	25.98	6.51
4	27.33	5.99
5	28.50	7.17

2.2 *The non-effect of phrasal length on peak height*

Figure 2 shows the schematized mean F0 contours of utterances of different lengths (from two to five pitch accents), holding the number of unstressed syllables between accents at three. All of the first peaks show a more or less constant pitch value (around 138 Hz), with the exception of the first peak of a two-accent utterance, which is significantly lower than the first peak of other phrasal lengths (around 115 Hz).

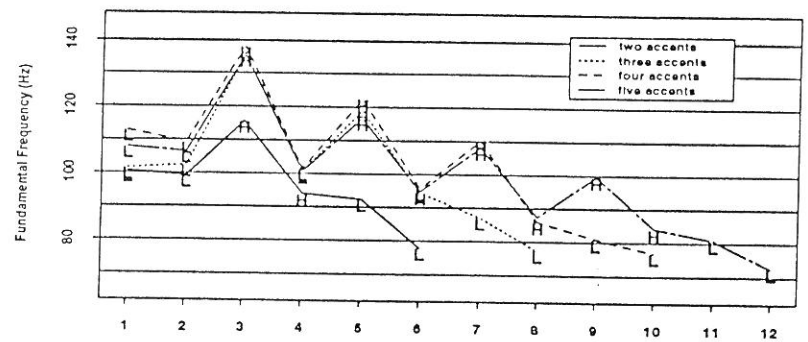

Figure 2: Schematized F0 contours of five utterances of different lengths (from 2 to 5 pitch accents), keeping constant (at 3) the number of intervening unstressed syllables between pitch accents.

Thus, in general, our data do not exhibit phrasal-length effects on the height of F0 peaks (leaving aside peaks of two-accent phrases, which were then not used in the data analyses). Rather, when holding the number of intervening unstressed syllables constant, the height of a given peak is determined by its position in the utterance, regardless of utterance length or the number of accents that follow. Therefore, phrasal length is not a factor that need be of concern to us as we examine whether or not the distance between pitch accents has an effect on peak height.

2.3 *The distance between pitch accents: an effect on peak height?*

We may now turn to the central question of this paper: is a time-dependent declination effect present in the descending F0 peak patterns produced by our speaker? If so, as stated earlier, given two adjacent peaks, the second one will show a progressively lower height as its distance from the previous peak increases. In this study, this distance is measured in two ways: 1) in terms of the number of unstressed syllables between accents, and 2) in terms of real time in milliseconds.

In regard to distance in terms of intervening unstressed syllables, Table 3 shows the mean difference between the F0 values of all adjacent peaks, grouped into utterances of different lengths and intervening unstressed syllable conditions. The data seem to show that the one-syllable condition triggers less peak decay, a result indicative of a time-dependent effect. However, notice that there is little difference between the two- and three-syllable conditions in the amount of peak decay present.

Table 3: Mean absolute peak difference in Hz between adjacent peaks with differing number of intervening unstressed syllables (1 to 3) in the position indicated.

Number of accents	Interv Sylls	Pos. 1-2	Pos. 2-3	Pos. 3-4
3	1	9.33		
	2	14.11		
	3	16.10		
4	1	7.62	10.03	
	2	14.55	14.80	
	3	14.93	15.60	
5	1	7.93	8.10	5.84
	2	15.46	13.25	8.82
	3	16.29	14.03	10.47

The different behavior of the one-syllable condition vs. the two- and three-syllable conditions cannot be due to a consistent and regular declination effect, since phrase-medial F0 peaks are not also consistently affected by having two as opposed to three unstressed syllables between them. How can we account, then, for the smaller degree of peak decay shown in the one-syllable condition, if this result is not attributable to a declination effect? The explanation might be that in our data, only this syllable condition creates a syntactic boundary separating two simple NPs. The presence of a syntactic boundary could cause a slight resetting of pitch, which is common in other languages. Thorsen (1980:21), for example, found this very effect in Danish. Further experimental data from Spanish would need to be studied in order to confirm or reject this proposed hypothesis.

Looking now at peak height itself in Figure 3 (as a supplement to the mean differences between adjacent peak heights, presented in Table 3), we plotted the mean absolute values of four successive F0 peaks, grouping utterances with one, two, or three unstressed syllables before each target peak. (To generate this figure, similarly positioned peaks preceded by like syllable conditions (e.g., all peaks in accent position two preceded by one syllable) were averaged over the three to five-accent utterance lengths, as deemed reasonable to do from Table 3.) In all but the fourth position, the one-syllable condition again results in a smaller degree of lowering of the following peak than the two or three-syllable conditions. T-tests comparing peaks in the one-syllable condition to peaks in the two or three-syllable conditions show that the differences are statistically significant ($p < 0.01$) in all cases but the fourth position in the utterance. Again, this one-syllable-condition result might point to a declination effect. However, if this were the case, just as the one-syllable condition provokes less lowering than the two-syllable one, we would expect the two-syllable condition to provoke less lowering than the three-syllable one. Yet the latter does not occur: the two- and three-syllable conditions do not bring about a different lowering effect, not making it possible to claim the presence of a declination effect.

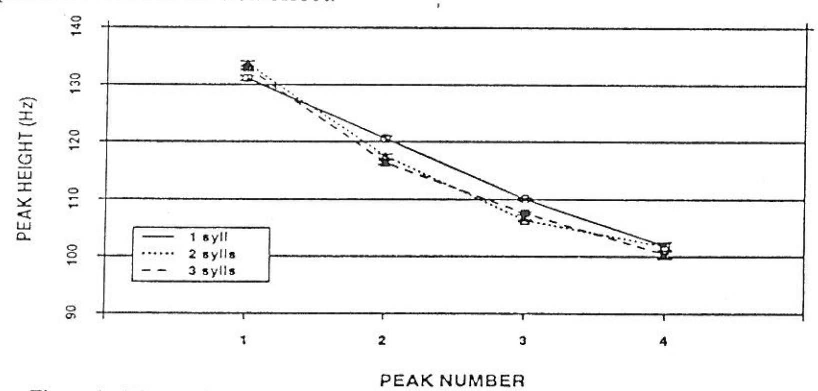

Figure 3: Mean adjacent peaks corresponding to an increase in the number of unstressed syllables between pitch accents (1 to 3). Solid lines link the peaks separated by 1 unstressed syllable, dotted lines link those separated by 2 unstressed syllables, and dashed lines those separated by 3 unstressed syllables.

The clearest evidence for the lack of a time-dependent declination effect on peak height can be seen by measuring the distance between the peaks in terms of real time rather than in terms of intervening unstressed syllables. We performed an analysis that checked for a correlation between the time interval between two peaks and the difference in Hz between them. While the complete list of results (or the 'correlation coefficients') of this analysis (performed within-groups) is too lengthy to include here (there are approximately 90 correlation coefficients per group), they show that the distance in time between two peaks is not correlated with the difference in Hz between them (i.e., there is not a positive correlation between these two variables). If anything, there is a very small tendency to increase the height of the second peak as its distance from the previous one increases, as the negative correlation coefficients obtained are highly significant (at $p < 0.01$).

Thus, our results, especially the one based on the more exact real time measurement, point to the absence of a time-dependent declination effect in the downstepped F0 contours of our Mexican speaker (as an increase in the distance between two adjacent F0 peaks does not bring about a consistent lowering of the second peak). This outcome constitutes evidence against the view that time-dependent declination is an always present, automatic pitch mechanism (see Lieberman 1967, Cooper & Sorensen 1981, Fujisaki 1983, 1988).

2.4 *Downstep patterns*

The fact that the F0 levels attained by peaks in a given position in the utterance are fairly constant (regardless of utterance length or distance between peaks) strongly indicates that F0 descent is governed solely by a ratio of decay between peaks in Mexican Spanish. Liberman & Pierrehumbert (1984:190) found that they could accurately predict the height of F0 peaks in a downstepping contour by reducing the previous peak by a fixed proportion. Even though these authors calculated the ratios in reference to an abstract reference line, we decided to use a non-abstract scaling line, i.e., the baseline of our speaker (Baseline=75 Hz). The value of each peak position was calculated by subtracting the fixed baseline value from the absolute height of the peak. The equation used to uncover the ratio of decay, or downstep ratio, is presented in (2):

(2) Downstep Ratio (r) = (P(x+1)-B) / (P(x)-B),
where P(x) = peak height of a peak in position x,
and B = baseline value (75 Hz).

Table 4 shows the mean absolute value of peaks in different phrasal positions (Peaks One through Four) for each phrasal length (two to five pitch accents), and shows in parentheses the downstep ratio between adjacent peaks (calculated using the formula in (2)). Since in our data F0 peaks separated by

one unstressed syllable behaved differently, we incorporated only the data points of utterances with two or three unstressed syllables between pitch accents in the target position to calculate these means.

Table 4: Mean absolute value of peaks (in Hz) and the ratio of decay between them (in parentheses) in phrases of different lengths, using only utterances with 2 or 3 unstressed syllables in the *target* position.

Number of accents	Peak 1	Peak 2	Peak 3	Peak 4
2	116.970			
3	131.248 (0.78)	119.385		
4	133.546 (0.78)	121.248 (0.70)	107.688	
5	134.154 (0.78)	121.707 (0.75)	110.377 (0.77)	102.295

The results shown in Table 4 indicate that the pattern of decay of peaks in a Spanish downstepping contour, as in English, is successfully described by applying a constant downstep ratio that iteratively reduces the value of each peak. Figure 4 plots the actual mean F0 peaks from Table 4 (solid lines) and compares them to the predicted F0 peaks (dotted lines) for utterances of different lengths.

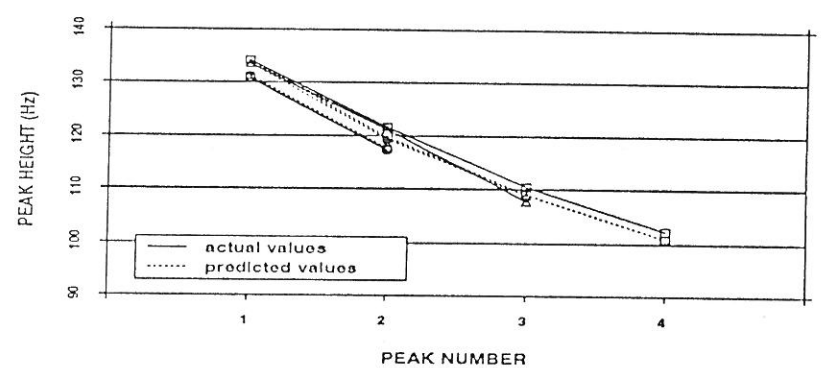

Figure 4: Actual and predicted F0 peak values in utterances of different lengths (from 3 to 5 pitch accents). The plotting symbol "o" is used for 3-accent utterances, "Δ" for 4-accent utterances, and " " for 5-accent utterances.

The predicted values were calculated using the mean initial peak for each of the phrasal lenghts as a starting point. Then, each peak was calculated as a constant fraction ($r = 0.76$) of the previous one, with reference to the baseline of our speaker (B), as shown in (3):

(3) $P(x+1) = r \cdot (P(x)-B) + B$,
where $P(x)$ = peak height of a peak in position x,
r = downstep factor (or ratio),
and B = baseline value (75 Hz).

The clear proximity of the actual and predicted F0 peak values is further evidence that in Spanish, a downstep ratio is sufficient to account for peak decay in descending contours, without recourse to a time-dependent declination function.

3.0 *Conclusion*

In our data, F0 peak height is predicted quite successfully by the application of a local 'downstep' ratio, or constant reduction in the previous peak's F0 value, relative to the baseline of the speaker. In other words, the descending peak pattern found in Mexican Spanish can be modelled exclusively by an exponential decay to a constant nonzero asymptote (the baseline of the speaker). This produces contours with a sharper degree of decay between the first and second peaks and a progressively lower degree of decay over the phrase-medial ones.

Distance between pitch accents (measured both in time and in syllables) has no real effect on F0 peak height. This clearly indicates that F0 descent in our Mexican Spanish contours cannot be attributed to a time-dependent descending F0 component and that, therefore, declination cannot be held as a non-suppressible, or automatic, mechanism in F0 control (see Lieberman 1967, Cooper & Sorensen 1981, Fujisaki 1983, 1988).

Although more work is needed to test the generality of such results for greater distances between peaks, or for a contour containing a series of non-downstepped H* accents, we hope to have contributed to an understanding of one of the phonetic parameters controlling a type of simple declarative contour in Spanish. We are currently focusing on the prediction of other crucial parts of this contour type, such as the low ('valley') values (see Prieto 1995) and the 'final fall' (see Prieto, Shih & Nibert 1995). Finally, our results have revealed a consistent tendency to reset pitch across two nominal phrases. The interesting issue of the scaling of peaks across syntactic constituents deserves further investigation if we want to advance in the description and understanding of F0 contours of longer utterances.

REFERENCES

Cooper, William E. & John M. Sorensen. 1981. *Fundamental Frequency in Sentence Production*. Heidelberg: Springer.
Fujisaki, Hiroya. 1983. "Dynamic Characteristics of Voice Fundamental Frequency in Speech and Singing". *The Production of Speech*. ed. by P. F. MacNeilage, 39-55. New York & Berlin: Springer-Verlag.
----------. 1988. "A Note on the Physiological and Physical Basis for the Phrase and Accent Components in the Voice Fundamental Frequency Contour". *Vocal Physiology: voice production, mechanisms, and functions*. ed. by O. Fujimura, 347-355. New York: Raven.
Ladd, D. Robert. 1984. "Declination: a review and some hypotheses". *Phonology Yearbook*. 1.53-74.
Liberman, Mark & Janet Pierrehumbert. 1984. "Intonational Invariance Under Changes in Pitch Range and Length". *Language Sound Structure*. ed. by Mark Aronoff & Richard Oehrle, 157-233. Cambridge, Mass.: MIT Press.
Lieberman, Philip. 1967. *Intonation, Perception, and Language*. Cambridge, Mass.: MIT Press.
Navarro-Tomás, Tomás. 1944. *Manual de entonación española*. New York: Hispanic Institute in the United States.
Pierrehumbert, Janet. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT.
---------- & Mary Beckman. 1988. *Japanese Tone Structure*. Cambridge, Mass.: MIT Press.
Prieto, Pilar. 1995. "Register Shift in Downstepping Contours". Ms., AT&T Bell Laboratories, Murray, Hill, NJ.
----------, Chilin Shih & Holly Nibert. 1995. *Pitch Downtrend in Spanish*. Ms., AT&T Bell Laboratories, Murray Hill, NJ.
Sosa, Juan Manuel. 1991. *Fonética y fonología de la entonación del español hispano-americano*. Ph.D. dissertation, Univ. of Massachussets, Amherst.
Talkin, David. 1989. "Looking at Speech". *Speech Technology*. 11.384-400.
Thorsen, Nina. 1980. "Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in ASC Danish". *Annual Report of the Institute of Phonetics, University of Copenhagen* 14.1-29.