Effects of Phrasal Length and Time Distance between Peaks on Peak Height in Mexican Spanish

*Pilar Prieto, Holly Nibert *and Chilin Shih AT&T Bell Laboratories*

Abstract

A speaker of Mexican Spanish read a total of 809 declarative utterances of varying length (from 2 to 5 pitch accents) and of varying distance between pitch accents (from 1 to 3 intervening unstressed syllables). The resulting contour can be described as a series of simple peaks or H* pitch accents, following [1], where each peak is lower than the one preceding it. An analysis of the data indicates that neither phrasal length nor distance between adjacent pitch accents have an effect on peak height. Rather, the height of a given peak is determined by its position within the utterance and can be predicted quite successfully by reducing the previous peak's pitch value by a constant *downstep ratio*.

INTRODUCTION

The purpose of the present experiment is to examine two controversial questions regarding the behavior of pitch downtrend (i.e., the general lowering of pitch over the course of a phrase) in Mexican Spanish: a) what is the effect of time distance between peaks on peak height? b) what is the effect of phrasal length on peak height?

The first question concerns the nature of downtrend: is it governed by a gradual time-dependent *declination* effect, by a ratio- driven *downstep* effect, or by both? The former type of effect, which has been claimed to occur in many languages (e.g., [2] for Japanese, [3] for Danish, [4] for English, among many others), predicts that the *length of time interval* between two peaks in a fundamental frequency contour will affect the pitch level of the second peak: the greater the distance between the two peaks, the lower the F_0 value of the second peak. Yet, studies like [5] found that adjacent F_0 peaks in English descending contours displayed an *invariant* ratio of decay, regardless of the distance between them. In our experiment, we will examine whether the time-dependent declination effect is a necessary component in pitch downtrend modelling.

The second issue investigated by the present experiment is the effect of phrasal length (measured by the number of pitch accents in an utterance) on the scaling of initial peaks. Recent instrumental studies make contradictory claims about the relationship between the length of an utterance and the height of the phrase-initial peak. While authors like [6] report a significant increase in the height of the first peak in longer utterances, other authors find that peak values are more or less constant in a given position, regardless of utterance length, see [5], [7], [8], [9].

*University of Illinois at Urbana-Champaign

EXPERIMENTAL DESIGN

The database consisted of 120 declarative *listing* phrases, all comprised of a noun phrase modified by an increasing number of embedded noun or prepositional phrases. They were obtained by exhaustively combining phrases with 2 to 5 pitch accents with 1 to 3 intervening unstressed syllables. The three following utterances correspond to the 2-accent group (2 accents with 1, 2, and 3 unstressed syllables in between): 1) *ráyo lúna*; 2) *ráyo de lúna*; 3) *ráyo de la lúna*. To facilitate the comparison between peaks in the same position, the ordering of the content words has been kept constant in all of the utterances. Thus, the structural characteristics of the database make possible a strict analysis of both the distance between pitch accents and phrasal length on pitch height.

A male speaker of Mexican Spanish read the 120 utterances at least five times in random order, for a total of 809 sentences. The speaker was trained to read the sentences at a normal speech rate using a descending pitch pattern. Figure 1 illustrates a typical utterance with five pitch accents together with our labeling scheme. In accordance with [1] the exhibited contour could be characterized as a series of simple peaks, or H* accents. Each peak is downstepped and is lower than the one preceding it.

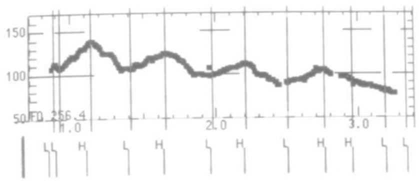

Figure 1: Five-accent utterance

F_0 peaks and valleys of each utterance were manually labelled using the Waves speech analysis package of Entropic Inc. The points labeled H in Figure 1 (the highest absolute F_0 values for each accent) constitute the target data in the present study. The utterance-final pitch accent was excluded from the analysis, since it did not display an F_0 peak in our data. The contour in Figure 1 is a 5 pitch-accent sentence with only four clear peaks.

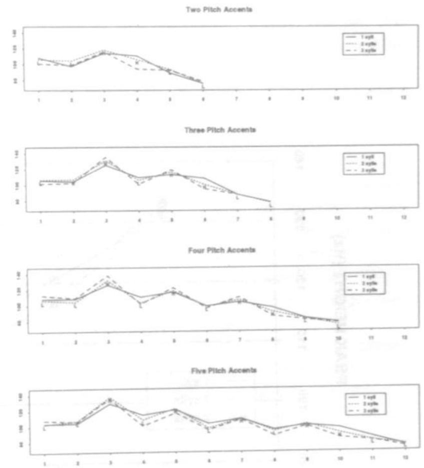

Figure 2: Schematized F_0 contours

RESULTS

Before the analysis, we checked that the utterances were produced over a more or less constant pitch range, since pitch range variation is an additional factor that could confound the effects of the factors under study. In our data, pitch range (measured as the distance in Hz from the lowest F_0 point to the following peak) was rather stable for a given phrasal position (which, of course, decreased steadily over the course of an utterance).

Effects of Phrasal Length

Figure 2 shows the schematized mean F_0 contours of utterances of different lengths (from 2 to 5 pitch accents). The height of the first peak is nearly constant in the 3 to 5-accent cases (around 138 Hz). The first peak of a 2-accent utterance, however, is much lower. For each utterance length, there are a set of three mean contours that show a steady increase in the number of unstressed syllables between accents (1 to 3). Solid lines represent a distance of 1 unstressed syllable between accents, dotted lines correspond to 2 unstressed syllables, and dashed lines show 3 unstressed syllables. The peak heights are not affected by the varying number of intervening unstressed syllables.

In general, our data do not exhibit phrasal length effects on peak height, leaving aside the first peak of 2-accent utterances (which is not included in further data analyses). Therefore, utterance length need not be of concern when we analyze the effects of the distance between pitch accents.

Effects of Distance between Peaks

Is a time-dependent declination effect present in the descending F_0 peaks produced by our speaker? If so, we would expect the second one of two adjacent peaks to be lower as the distance between the two increases. We measured this distance in two ways: in terms of number of unstressed syllables between accents (from 1 to 3), and in terms of real time.

Table 3 shows the mean distance in Hz between adjacent peaks, grouped into utterances of different lengths and number of intervening unstressed syllables. The results show that, for a given phrasal position, the distance in Hz between adjacent accents is nearly the same, except in one case: the data seem to show that the one-syllable condition triggers less peak decay. This pattern is also observed in Figure 2, which plots the mean absolute values of four successive F_0 peaks, using utterances with one, two, or three unstressed syllables before the target peak. T-tests comparing the peaks in the three conditions (1, 2, 3 preceding unstressed syllables) show that while peaks in the two and three-syllable conditions belong to the same group, peaks in the one-syllable condition are significantly different (at $p. < 0.01$).

We claim that the effect triggered by the

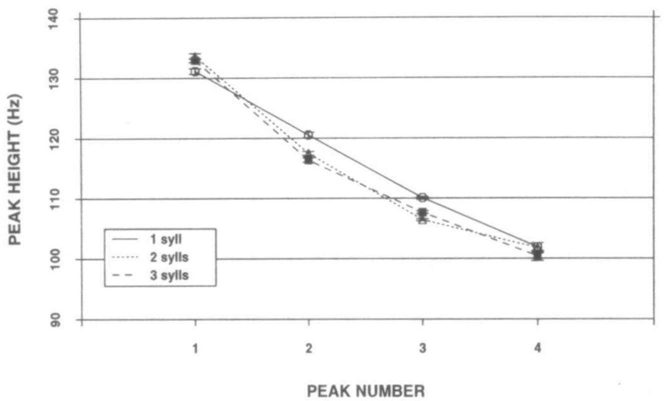

Figure 3: Mean adjacent peaks corresponding to an increase in the number (1 to 3) of unstressed syllables between pitch accents. Solid lines link the peaks separated by 1 unstressed syllable, dotted lines link those separated by 2 unstressed syllables, and dashed lines link those separated by 3 unstressed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

Table 1: **Mean absolute peak didfference 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

one-syllable condition is attributable to a slight pitch reset created by a syntactic boundary separating two simple NPs (the one-syllable condition is the only one with this syntactic configuration). This phenomenon has been reported for other languages (see [3] Danish), and we will leave this issue for further investigation.

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. We performed a within-group correlation analysis between the time interval (in ms) between two peaks and their difference in Hz. Surprisingly, if anything, there was a small tendency to increase the height of the second peak as its distance to the previous peak increases (the negative correlation coefficients were highly significant (at $p < 0.01$)).

CONCLUSION

The analysis of the data indicates that neither phrasal length nor distance between adjacent pitch accents have an effect on peak height in the downstepped F_0 contours produced by our Mexican speaker. Rather, the height of a peak is determined by its position within the utterance: for example, the second H peak in a sequence will show a more or less constant pitch value, regardless of whether it appears in a three, four, or five pitch accent sentence or whether it is separated from the preceding accent by 1, 2, or 3 unstressed syllables. Further calculations show that peak height is predicted quite successfully by a constant reduction (also called *downstep ratio*) of the previous peak's pitch value. Thus, our results point to a lack of time-dependent declination effect and provide evidence against the view that global declination is an automatic and universally pitch mechanism [2], [6], [10], [11].

REFERENCES

[1] Pierrehumbert, J. (1980) *The Phonology and Phonetics of English Intonation.* Ph. D. Dissertation, MIT.

[2] Fujisaki, H. (1983) Dynamic Characteristics of Voice Fundamental Frequency in Speech and Singing. In *The Production of Speech* (P.F. MacNeilage, editor), pp. 39-55. New York and Berlin: Springer-Verlag.

[3] Thorsen, N. (1980) Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Lengt[illegible] ASC Danish. *Annual Report of the Instit[illegible] of Phonetics*, University of Copenhagen 1[illegible] 1-29.

[4] Liberman, M. (197[illegible]) The Intonational System of English. Ph. D. Dissertation, MIT. Cambridge, Massachussets.

[5] Liberman, M. & Pierrehumbert, J. (1984) Intonational invariance under changes in pitch range and length. In *Language Sound Structure* (M. Aronoff & R. Oehrle, editors), pp. 157-233. Cambridge: MIT Press.

[6] Cooper, W. E. & Sorensen, J.M. (1981) *Fundamental frequency in sentence production.* Heidelberg: Springer.

[7] Thorsen, N. (1981) Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in ASC Danish. — Supplementary Data. *Annual Report of the Institute of Phonetics*, University of Copenhagen 15, 13-47.

[8] Sternberg, S., Wright, C.E., Knoll, R.L. & Monsell, S. (1980) Motor programs in rapid speech: additional evidence. In *Perception and production in fluent speech* (R.A. Cole editor), pp 507-534. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

[9] van den Berg, R., Gussenhoven, C. and Rietveld, T. 1992. Downstep in Dutch: Implications for a model. In *Papers in Laboratory Phonology II: Segment, Gestures, Tone.* (G.J. Docherty and D.R. Ladd, editors), pp. 335-367. Cambridge: Cambridge University Press.

[10] Lieberman, P. (1967) *Intonation, perception, and language.* Cambridge: MIT Press.

[11] Fujisaki, H. (1988) A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In *Vocal Physiology: voice production, mechanisms and functions* (O. Fujimura, editor), pp. 347-355. New York: Raven.