Peak scaling in Greek and the role of declination

Amalia Arvaniti
Department of Linguistics, University of California, San Diego, 9500 Gilman Drive #0108,
La Jolla, CA, 92093-0108, USA
amalia@ling.ucsd.edu

ABSTRACT

The controversial role of declination in the scaling of accentual peaks was tested using Greek materials, consisting of sentences of increasing length (from two to five accented words) in which successive accents were separated by either two or four unstressed syllables. It was hypothesized that if peak scaling is locally controlled (i.e. independent of declination), the peaks in the long sentences would not be scaled lower than peaks found in the same position in the short sentences. The results showed that peaks were scaled lower in the long sentences, but the effect was very small. On the other hand, multiple regression showed that the best predictor of peak scaling was the scaling of the preceding peak, while other factors played a very small role. These results largely agree with results from Spanish and English and support the view that declination does not play a major part in peak scaling.

1. INTRODUCTION

The role of declination in the scaling of accentual peaks has been strongly disputed in intonational research. Results from English [1] and Spanish [2] suggest that the scaling of a given peak can be successfully modeled as a fraction of the scaling of the peak preceding it, i.e. without recourse to declination. Other researchers (e.g. [3], [4]), however, attribute peak scaling primarily to declination. Although this issue has important ramifications for intonation modeling, very few experimental studies so far have addressed the role of declination in peak scaling .

In this paper, declination is examined using data from Greek, in which it is possible to create relatively long declaratives uttered without phrase breaks. In such cases, the melody involves a series of identical prenuclear accents (L*+H in autosegmental/metrical notation; see [5] and [6]) on all content words (content words are rarely unaccented in Greek; [5]). These accents are typically followed by a H*+L nuclear accent, if focus is broad, and L-L% phrasal tones (see [5]). It was hypothesized that if peak scaling depends on declination, then in sentences with the same number of peaks, peaks that were closer together would be scaled higher than peaks in the same position that are separated by more segmental material. On the other hand, if peak scaling is not affected by declination the temporal distance between peaks would not affect their scaling.

2. METHOD

Materials: Sentences with two, three, four and five content words were created (see below). In half of the sentences, the distance between any two accents was two unstressed syllables (henceforth the *short sentences*). In the other half, this distance was increased to four unstressed syllables (henceforth the *long sentences*).

SHORT SENTENCES
2 words: [imeˈlina ciˈlina] “Melina and Lina”
3 words: [imeˈlina θarˈθi metiˈlina]
“Melina will come with Lina”
4 words: [imeˈlina θarˈθi me tiˈlina stiˈlimno]
“Melina will come with Lina to Lemnos”
5 words: [imeˈlina θarˈθi metiˈlina stiˈlimno tiˈdriti]
“Melina will come with Lina to Lemnos on Tuesday”

LONG SENTENCES
2 words: [imeˈlina metiboˈlina] “Melina with Paulina”
3 words: [imeˈlinamas θaʝiˈrisi metiboˈlina]
“Our Melina will return with Paulina”
4 words: [imeˈlinamas θaʝiˈrisi metiboˈlina stisaloˈnici]
“Our Melina will return with Paulina to Salonica”
5 words: [imeˈlinamas θaʝiˈrisi metiboˈlina stisaloˈnici mesolonˈðinu] “Our Melina will return with Paulina to Salonica via London”

The sentences were designed to be of similar structure; in addition, an effort was made to minimize the use of voiceless obstruents so as to obtain relatively smooth F0 contours. Each test sentence was preceded by a question that provided a context appropriate for broad focus, in order to make the reading of the sentences as natural as possible, and to avoid emphasis. For example, the question preceding the short two-word sentence was “Who else is coming?”; the long four-word sentence was preceded by “Where is Melina going after Paris?”.

Speakers: There were three speakers, two females in their thirties (GRF1 and GEF2) and one 22 year old male (GRM3). GRF1 and GRM1 are speakers of Athenian Greek, while GRF2 is from Thessaloniki.

Procedure: GRF1 and GRF2 were recorded on DAT tape in the recording studio of the Linguistics Department of

UCSD. The data of GRF1 were later re-digitized at 44.1 KHz using Pitchworks; GRF2's data were re-digitized using CoolEdit at 48 KHz. Finally, GRM1 was recorded in the same location but directly to disk, using an A/D converter, at a sampling rate of 44.1 KHz.

The sentences were typed on flashcards, each showing a question and answer set, and were read in random order, eight times by GRF1 and GRF2, and nine times by GRM1. None of the speakers read the questions. The materials were interspersed with materials for another experiment (see [7]). During the recording, the speakers were asked to repeat a sentence in the rare instances in which it was disfluent, had extra emphasis on a word or a phrasal boundary at some point. All repetitions were used for measurement.

Measurements: The following measurements, illustrated in Figure 1, were made using PRAAT, by simultaneous inspection of waveforms, spectrograms and pitch tracks.

- Initial Low (IL): the lowest F0 value early in the utterance. As can be seen in Figure 1, IL did not always coincide with the very beginning of the utterance; in fact IL usually co-occurred with the second syllable.
- Final Low (FL): the final and lowest F0 value in each utterance. This point was occasionally difficult to measure due to creak, but such instances were rare.
- The F0 of all peaks (H1, H2, H3, H4 and H5). Each peak was measured at the highest point of a rise in the vicinity of the accented syllable (see [6] on the late alignment of accentual peaks in Greek). If the rise ended in a short plateau, the first of a series of identical values was measured.

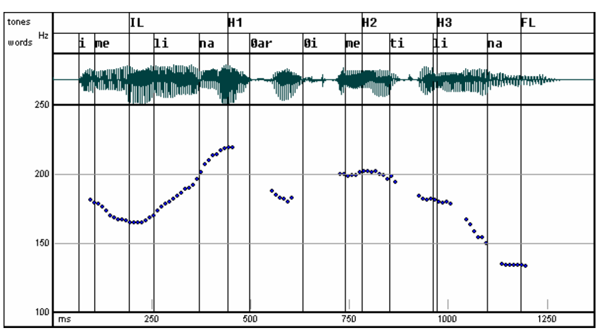

Figure 1: Waveform and pitch contour of the three-word short sentence as realized by GRF2; in the "words" tier, vertical lines correspond to syllable boundaries; in the "tones" tier, they correspond to F0 measurements.

3. RESULTS

To begin with, it was essential to establish that the speakers used a relatively constant pitch range across repetitions. In order to do so, the difference in Hz was calculated between H1 (the first peak in each utterance and highest F0 value) and FL (the final and lowest F0 value). Results are presented in Table 1, together with the values for IL, FL and H1. All these measurements suggest that pitch range was relatively stable for each speaker.

Speaker	IL Mean	IL s.d.	H1 Mean	H1 s.d.	FL Mean	FL s.d.	Range Mean	Range s.d.
GRF1	182	4	235	9	159	8	76	11
GRF2	172	9	240	17	139	6	101	17
GRM1	99	3	127	5	95	6	32	7

Table 1: Means and standard deviations of IL, H1, FL, and pitch range, for each speaker separately.

A two-way between-subjects ANOVA (SPEAKER × SENTENCE LENGTH) was run to see if peaks were scaled lower in long sentences. The results showed that this was the case ($F(1,694)=9.3$, $p<0.002$). However, the difference between peaks in short and long sentences was only 4 Hz (3 Hz for GRF1, 5 Hz for GRF2, and 4 Hz for GRM1). Thus, although SENTENCE LENGTH did play a part in peak scaling the effect was very small.

Similar results were obtained when the data were broken down by accent. The scaling of H1, H2, H3 and H4 was investigated using three-way ANOVAs (SPEAKER × SENTENCE LENGTH × NO OF WORDS), and that of H5 using a two-way ANOVA (SPEAKER × SENTENCE LENGTH). The results for H1 and H3 were non significant, but H2, H4 and H5 did show an effect of SENTENCE LENGTH (for H2, $F(1,176)=7.8$, $p<0.005$; for H4, $F(1,88)=18.3$, $p<0.0001$; for H5, $F1(1,44)=10.2$, $p<0.002$). As shown in Figure 2, this effect was similar to that observed in the pooled data, i.e. lower peak scaling in the long sentences; again the differences in peak height were very small, on average 3 Hz for H2, 9 Hz for H4, and 7 Hz for H5.

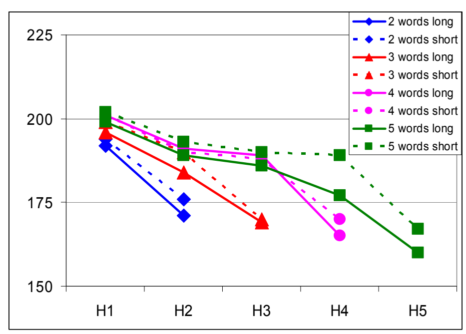

Figure 2: Mean scaling of the five peaks as a function of SENTENCE LENGTH and NO OF WORDS.

In addition, the ANOVAs on H1, H2, H3 and H4 showed an effect of NO OF WORDS, illustrated in Figure 3 (for H1, $F(3,176)=6.2$, $p<0.001$; for H2, $F(3,176)=47.45$, $p<0.0001$; for H3, $F(2,132)=78.96$, $p<0.0001$; for H4, $F(1,88)=67.4$, $p<0.0001$). Planned comparisons across the levels of this factor showed that peaks in the sentences with the smallest number of words were lower than peaks in the same position but in sentences with more words. However, in all cases, NO OF WORDS interacted with SPEAKER (for H1, $F(6,176)=2.16$, $p<0.04$; for H2, $F(6,176)=7.7$, $p<0.0001$;

for H3, F(4,132)=11.64, p<0.0001; for H4, F(2,88)=15.64, p<0.0001). These interactions were investigated using the Tukey HSD *post-hoc* test, which showed that the general effect described above was valid for H1 in the data of GRF1 only, and for H2, H3 and H4 in the data of GRF1 and GRF2 (for H1 two- vs. three- words, p<0.02; p<0.01 for all other significant comparisons). There was no effect for GRM1.

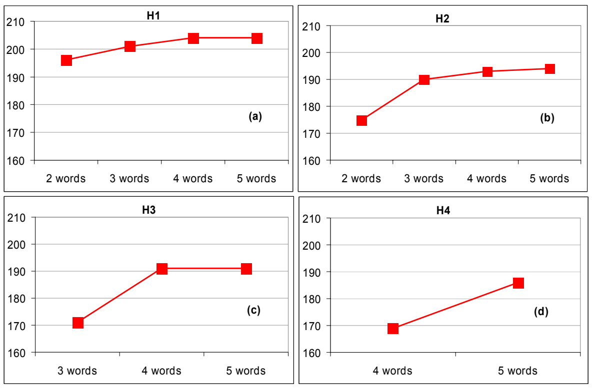

Figure 3: Means and standard errors for H1, H2, H3 and H4 as a function of NO OF WORDS.

Multiple regression models were developed on the basis of the ANOVA results reported above. Peak F0 was the dependent variable, while the predictors were added using a step-wise approach (see [8]). Models for the pooled data and for each speaker separately were created. The final peaks were included in the models, because models from which they were excluded fared much worse. The predictors, in descending order of correlation with the dependent variable, were PEAK ORDER, NO OF WORDS, and SENTENCE LENGTH. The four models are as follows:

Peak Height (pooled data): -202.79 – .25 PEAK ORDER + .12 NO OF WORDS + .039 SENTENCE LENGTH

Peak Height (GRF1): -32.84 – .94 PEAK ORDER + .471 NO OF WORDS + .078 SENTENCE LENGTH

Peak Height (GRF2): -356.11 – .79 PEAK ORDER + .382 NO OF WORDS + .118 SENTENCE LENGTH

Peak Height (GRM1): -217.58 – .61 PEAK ORDER + .309 SENTENCE LENGTH + .248 NO OF WORDS

As can be seen, PEAK ORDER was overall the best predictor. SENTENCE LENGTH was the worst predictor for GRF1 and GRF2 (and non significant when the data were pooled), but the second best predictor for GRM1. However, in all cases, the predictive value of the models was rather low: it was 74% for GRF1, 52% for the pooled data and GRF2, and 40% for GRM1. In other words, only the model for GRF1 could be considered satisfactory.

One possible explanation for the small effects observed in the ANOVAs and the low predictive power of the multiple regression models could be that peak scaling was more strongly affected by the actual duration of the unstressed syllables between peaks. This is particularly important to test in a language like Greek in which many unstressed vowels are elided (an effect particularly noticeable in the data of GRM1), so that syllable count may not accurately reflect duration. In order to test what the effect of actual duration was on the scaling of peaks, correlations were run for the difference in scaling between successive peaks and the duration of the interval separating them. The pooled data show a small (though significant) negative correlation (r = -0.3), suggesting that the longer the interval between peaks, the smaller their difference in scaling was (see Figure 4). Although this result is difficult to interpret, it neither points towards a powerful declination effect, nor is it strong enough to be a good predictor of peak scaling.

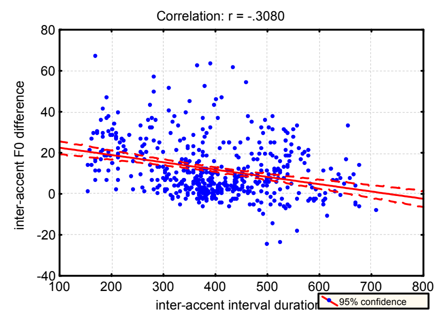

Figure 4: Difference in scaling between successive peaks as a function of inter-accent interval duration; pooled data.

An alternative, proposed by [1] and [2], is that the height of a given peak depends largely on the height of the peak preceding it. In order to see if this holds for Greek, as in English and Spanish, I tested the correlation of preceding peak height with following peak height. As can be seen in Figure 5, this correlation is very strong.

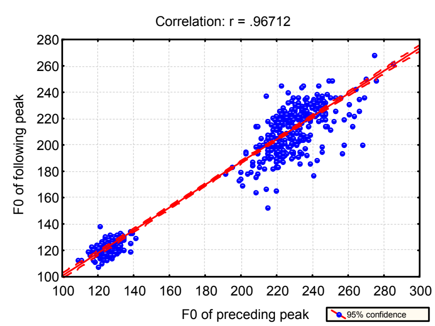

Figure 5: Peak height as a function of the height of the preceding peak.

Given this high correlation, another multiple regression model was created using the F0 of preceding peaks as the sole predictor of the F0 of following peaks.

Peak Height (pooled data): 14.22 + .967 preceding peak height.

As can be seen, this model surpasses the others by far: it is much simpler and predicts 93% of the variance. Adding the other predictors improves its performance somewhat (by about 2%), but the difference is very small. This model strongly suggests that the best predictor of the height of a peak is the height of the preceding peak.

4. DISCUSSION AND CONCLUSION

An experiment was devised to test the extent to which peak scaling in Greek is dependent on declination. It was hypothesized that if declination is crucial for peak scaling, peaks in long sentences would be scaled lower than peaks in a similar position (in terms of peak order) in short sentences. This hypothesis turned out to be supported by the data, although the magnitude of the observed effect was very small (4 Hz on average).

Similarly small effects on scaling were observed when the number of words in a sentence was considered. For example, Figure 3a shows that H1 is scaled lower in two-word sentences than in longer sentences. This connection between the scaling of the first peak and the length of the utterance has been reported before (e.g. [9]), and indirectly supports the importance of declination in peak scaling, since it shows that the top of the pitch range is set higher in preparation for longer utterances. However, as Figure 3a suggests, this effect is rather small in the present data. Compare this to the final lowering effect, observed in the other panels of Figure 3, which show that final peaks are scaled much lower than penultimate peaks in the same position (for a fuller discussion of final lowering see [7]).

On the other hand, the data showed a small negative correlation between the inter-accent interval duration and the F0 difference between successive peaks. As mentioned, this correlation, which was also observed in Spanish [2], is difficult to interpret. One possible explanation could be that the speakers reset their level if an accent was too far apart from the one preceding it. However, this does not seem plausible: the recordings were carefully monitored to avoid phrase breaks that could lead to such reset; moreover, reset would have been obvious during analysis and the data would have been discarded. As [2] also point out, however, it is certain that this negative correlation does not point towards a strong declination effect on peak scaling: if declination was important, a positive correlation would have been found instead.

Finally, multiple regression models were created to test the relative contribution of the various factors that according to ANOVA and correlation results affected peak scaling. One interesting point regarding these models is that the exclusion of final peaks worsened the models' performance. The reasons for this are not entirely clear, since final lowering did take place in the data (see [7]). It is possible, however, that the effect in Greek was not as pronounced as in other languages, such as Spanish (cf. data in [2] and present Figure 2); this difference may have allowed for modeling that included the final peaks. More to the point, models that took into account factors such as the length of utterances, the position of the peak and the number of words in the utterance left about half of the variance unexplained (a result also reported by [2] for Spanish). In contrast, a much simpler model in which the sole predictor of peak scaling was the scaling of the preceding peak accounted for 93% of the variance. Although it is clear that both a more detailed model for the present data and research involving more speakers are necessary, the present results clearly suggest that the role of declination in peak scaling is indeed small. Peak scaling in Greek, as in Spanish and English, can be satisfactorily modeled on the basis of preceding peak height alone.

ACKNOWLEDGEMENTS

The research reported here was supported by a UCSD COR grant (no RB170C-ARVANITI) to Amalia Arvaniti. I would like to thank Svetlana Godjevac for carrying out the measurements, and Pauline Welby and Chris Barker for their invaluable help with the intricacies of PRAAT.

REFERENCES

[1] M. Liberman and J. Pierrehumbert, "Intonational invariance under changes in pitch range and length," in *Language Sound Structure*, M. Aronoff and R. T. Oehrle, Eds., pp. 157–233. Cambridge MA: The MIT Press, 1984.

[2] P. Prieto, C. Shih, and H. Nibert, "Pitch downtrend in Spanish," *Journal of Phonetics*, vol. 24, pp. 445-473, 1996.

[3] H. Fujisaki, "Dynamic characteristics of voice fundamental frequency of speech and singing," in *The production of speech*, P. M. MacNeilage, Ed., pp. 39-45. New York: Springer Verlag, 1983.

[4] N. Thorsen, "Intonation and text in Standard Danish," *Journal of the Acoustical Society of America*, vol. 77, pp. 1205-1216.

[5] A. Arvaniti and M. Baltazani (in press), "Intonational analysis and prosodic annotation of Greek spoken corpora," in *Prosodic Models and Transcription: Towards Prosodic Typology*, S. Jun, Ed. Oxford, UK: Oxford University Press.

[6] A. Arvaniti, D. R. Ladd and I. Mennen, "Stability of tonal alignment: the case of Greek prenuclear accents," *Journal of Phonetics*, vol. 26, pp. 3-5.

[7] A. Arvaniti and S. Godjevac (this volume)

[8] T. Rietveld and R. van Hout, *Statistical Techniques for the Study of Language and Language Behaviour*, Berlin and New York: Mouton de Gruyter, 1993.

[9] W. F. Cooper and J. M. Sorensen, *Fundamental frequency in sentence production*, Heidelberg: Springer, 1981.