



## PROSODIC DIFFERENCES IN READING STYLE: ISOLATED vs. CONTEXTUALIZED SENTENCES

Juan M. Garrido, Joaquim Llisterra, Carme de la Mota and Antonio Ríos

*Departament de Filologia Espanyola, Universitat Autònoma de Barcelona, Edifici B, 08193 Bellaterra, Barcelona, Spain.  
Fax (34.3) 581.16.86. E-mail: Joaquim.Llisterra@cc.uab.es*

### ABSTRACT

*Global properties of the Fo contour and local phenomena related to lexical stress have been analyzed in a set of read isolated sentences. The results have been compared with those obtained for the same set of sentences embedded in paragraphs. The results show that the main differences are concentrated on global Fo contours.*

**Keywords:** Prosody, Reading Style, Declination, Stress.

### 1. INTRODUCTION

It has been pointed out in the literature that isolated sentence intonation is modified when sentences are part of a larger textual unit. After Umeda's pioneering work [1], Lehiste [2], Grønnum Thorsen [3,4] and Ladd [5], among others, have contrasted the prosodic properties of isolated sentences with the characteristics of sentences embedded in paragraphs, to conclude that individual sentence intonation contours are modified by the textual context in which they appear. The idea of a "superordinate textual information structure" has been particularly proposed by Grønnum Thorsen [3].

The purpose of the present study is to provide additional data for Spanish showing which prosodic parameters are modified when sentences are read in isolation and when they are contextualized in paragraphs. This should help to improve the prosodic rules currently used in text-to-speech conversion.

### 2. PROCEDURE

The corpus consists of five different sentences read in two conditions: isolated and contextualized within a text. Each contextualized sentence appears in the text in three positions: initial, medial and final.

In order to avoid detection errors in the pitch tracking algorithm each sentence is formed only by voiced segments. The number of syllables (15) is kept constant for each sentence. The rhythmical pattern consists of an anacrusis followed by four dactyls and a final trochaic foot. The syntactic pattern is also common to all sentences: a subject NP complemented by a PP, and a transitive verb with a modified internal argument.

Five different paragraphs have been constructed. In each paragraph the same sentence appears in the three positions mentioned above. Four filler sentences have been added in order to obtain a semantically coherent paragraph. The first

and the third fillers are simple sentences and have about 15 syllables. The second and the fourth are complex sentences linked by connectors and have about 30 syllables. So, the target sentences in medial and final position are always preceded by a long sentence. All the occurrences of the target sentence contain pragmatically relevant information; their contextual effect reinforces the meaning of the preceding or following utterances. The texts were presented to the subjects in random order. The sentences to be read in isolation were included in a list with other filler sentences and were read before the texts.

The corpus was read by five male speakers of standard Peninsular Spanish, aged between 25 and 30, students at the UAB. One hundred target sentences were thus obtained.

The recording sessions took place in a sound treated room at the Phonetics Laboratory of the UAB using a Tascam 112 cassette recorder and a Sennheiser MKH20 microphone.

The signal was low-pass filtered and digitized at 10 KHz. sampling rate using a MacAdios II™ card. Waveform displays, broad-band spectrograms and Fo contours were plotted for each target sentence using the Mac Speech Lab II™ software package running on an Apple Macintosh IICx. Fo contours were obtained with a pitch tracking algorithm using an auto-correlation technique.

The following measurements were taken: Fo and duration values of stressed syllables, Fo and duration values of unstressed syllables showing an Fo maximum and Fo values of the syllables occurring at minima in the Fo contour. The Fo values for each syllable have been considered at the turning point of the syllable contour, if any, or at the mid-point of the syllable nucleus. Fo maxima and minima are defined in relation to Fo values found both in the preceding and following syllables. A maximum (P) is established if the Fo values in the preceding and following syllable are both lower than the Fo value found for the measured syllable. Correspondingly, a minimum (V) is detected when Fo values in the preceding and the following syllable are both higher than the Fo value in the measured syllable. In initial and final sentence syllables only the Fo value of the following or the preceding syllable is considered.

Mean and standard deviation for Fo and duration values within each condition have been calculated. Significant differences between the conditions have been assessed by means of one-factor ANOVA tests. Student t-tests have also been used to compare differences between groups.

### 3. RESULTS

#### 3.1. Global shape of Fo contours

##### 3.1.1. Declination

Top and bottom declination lines have been calculated for each sentence. Top declination lines link Fo values from the first to the last maximum found in the contour ( $P_I$  and  $P_F$ ); bottom declination lines link the first and the last minimum in the contour ( $V_I$  and  $V_F$ ). Values corresponding to the difference between Fo values at  $P_I$  and  $P_F$  on the one hand, and between  $V_I$  and  $V_F$  on the other, have been obtained. Thus, it was possible to calculate the rate of decay per syllable in the whole sentence for each condition (see Table 1).

	Isolated	T.initial	T.medial	T.final
topline	4.91 (0.98)	5.47 (1.43)	2.97 (2.74)	2.44 (1.74)
bottomline	1.26 (0.62)	1.64 (1.30)	0.82 (1.01)	0.77 (0.86)

Table 1: Mean value in Hz/syll. and s.d. of the rate of decay per syllable in the top and bottom lines

When the topline is observed, significant differences are found between a group formed by isolated and text initial sentences and another group formed by text medial and text final sentences. If the bottomline is considered, the only significant differences are found between text initial and text medial and between text initial and text final sentences.

Differences in Fo values between  $P_I$  and  $V_I$  and between  $P_F$  and  $V_F$  have also been calculated for each sentence. This gives information about the difference between the top and bottom declination lines at the beginning and at the end of the sentence (see Table 2).

	Isolated	T.initial	T.medial	T.final
$P_I - V_I$	51.08 (8.69)	52.84 (10.16)	38.44 (12.96)	29.12 (9.76)
$P_F - V_F$	22.88 (12.28)	24.60 (12.83)	20.44 (17.44)	15.80 (12.99)

Table 2: Mean values in Hz. and s.d. corresponding to the differences between  $P_I$  and  $V_I$  and between  $P_F$  and  $V_F$ .

When the beginning of the sentences is considered, significant differences are found among all conditions except between isolated and text initial sentences. When the end of the sentences is examined, significant differences are found only between text final sentences and the group formed by isolated and text initial sentences.

Differences among conditions related either to  $P_I$  or to  $P_F$  have also been observed (see Table 3).

	Isolated	T.initial	T.medial	T.final
$P_I$	165.32 (18.08)	168.68 (16.99)	143.44 (27.14)	132.72 (21.48)
$P_F$	119.80 (13.59)	119.04 (14.49)	113.80 (17.37)	108.64 (16.77)
$V_I$	96.92 (11.99)	115.84 (11.64)	105 (16.32)	103.60 (13.19)
$V_F$	96.92 (17.38)	94.44 (18.27)	93.36 (19.31)	92.84 (17.87)

Table 3: Mean values in Hz. and s.d. of  $P_I$ ,  $P_F$ ,  $V_I$  and  $V_F$ .

As far as  $P_I$  is concerned, significant differences are found between all conditions except between the isolated sentences and the text initial ones.  $P_F$  shows significant differences between isolated and text final and between text initial and text final sentences.

When the Fo measured at  $V_I$  is observed, significant differences are found between all conditions, except between isolated and text initial and between text medial and text final sentences. No significant differences appear among the conditions when the Fo of  $V_F$  is considered.

##### 3.1.2. Range

The range has been calculated by subtracting the value corresponding to the lowest minimum from the value corresponding to the highest maximum for each sentence. The results obtained are shown in Table 4:

	Isolated	T.initial	T.medial	T.final
range	69.72 (13.13)	76.20 (18.26)	55.68 (16.81)	45.52 (13.94)

Table 4: Mean value of range in Hz. and s.d..

Significant differences are found among all conditions except between isolated and text initial sentences.

#### 3.2. Prosodic features and lexical stress

##### 3.2.1. Fo maxima

Fo maxima have traditionally been considered a prosodic cue of lexical stress. In order to obtain a first approximation to the stress pattern of the sentences, the average number of maxima per sentence in each condition has been calculated: 4.28 for isolated, 4.16 for initial, 4.08 for medial and 4.12 for final sentences. It is clear that the average number of maxima per sentence does not change among conditions.

The distribution of Fo maxima has been compared with the

distribution of lexical stressed syllables along the sentence. Results are summarized in Table 5.

	Isolated	T.initial	T.medial	T.final
one syllable before	4.08% (4 cases)	2.10% (2 cases)	2.15% (2 cases)	2.94% (3 cases)
stressed syllable	18.36% (18 cases)	18.94% (28 cases)	22.58% (21 cases)	26.47% (27 cases)
one syllable after	77.55% (76 cases)	78.94% (75 cases)	75.26% (70 cases)	70.58% (72 cases)

Table 5: Percentage of occurrence and number of cases of Fo maxima relative to the stressed syllables.

It can be seen from the preceding distribution that more than 70% of Fo maxima appear one syllable after the stressed one. Approximately one fourth of the number of Fo maxima appears simultaneously with the stressed syllables. Differences between conditions do not seem to be significant, specially if the number of cases is taken into account.

The coincidence between stress and Fo maxima in the same syllable has been analyzed for each condition. Table 6 summarizes the results, based on 125 syllables per condition.

	Isolated	T.initial	T.medial	T.final
with P	20%	18.4%	22.4%	23.2%
without P	80%	81.6%	77.6%	76.8%

Table 6: Percentage of occurrence of stressed syllables showing or lacking Fo maxima (P).

It can be seen that in all conditions about 80% of the stressed syllables do not coincide with an Fo maximum. Differences among conditions are not found.

When considering the occurrence of Fo maxima in the corpus, it is observed that around 75% of the total number of maxima appear in unstressed syllables. The differences among conditions are shown in Table 7.

	Isolated	T.initial	T.medial	T.final
with P	76.6%	77.8%	72.5%	72.8%

Table 7: Percentage of Fo maxima (P) occurring at unstressed syllables.

### 3.2.2. Syllable duration

The duration of both stressed and unstressed syllables coinciding with Fo maxima has been considered in order to

compare changes in the four conditions.

Results for the duration of stressed syllables are shown in Table 8. No significant differences are found among the four conditions.

	Isolated	T.initial	T.medial	T.final
stressed syllables	163.46 (42.12)	155.48 (40.73)	157.08 (40.99)	155.96 (42.62)

Table 8: Mean duration in ms. and s.d. of stressed syllables.

Since not all stressed syllables show an Fo maximum, duration measurements are given separately for syllables associated with a maximum and for syllables which are not associated with a local maximum in the Fo contour.

	Isolated	T.initial	T.medial	T.final
stressed with P	178.28 (54.37)	181.26 (58.38)	175.57 (52.56)	161.03 (55.84)
stressed without P	159.76 (37.90)	149.66 (40.73)	151.74 (35.57)	154.43 (37.98)
unstressed with P	138.69 (54.23)	123.87 (37.95)	129.36 (42.58)	131.65 (45.25)

Table 9: Mean duration in ms. and s.d. of Fo maxima (P) in stressed and unstressed syllables.

No significant differences for stressed syllables showing an Fo maximum are found among the four conditions. The same result is found for stressed syllables without Fo maxima.

When the duration of unstressed syllables is compared to the duration of stressed syllables significant differences appear for all conditions.

A Pearson correlation coefficient has been calculated in order to assess the possible correlation between Fo values and syllable duration for each type of syllable. In all conditions, there is a clear lack of correlation between these two parameters.

## 4. DISCUSSION

The presence of a declination tendency has been detected in the four conditions analyzed. The top and bottom lines clearly show this trend (see Figure 1). It can also be seen that top and bottom lines are not parallel: the topline has a steeper slope than the bottomline in the four conditions, as has been reported in earlier studies. The presence of a declination line is constant across speakers and sentences; these results partially differ from those reported by Umeda [1].

In addition, the presence of a "paragraph declination" has also been found (see Figure 1). A paragraph declination line can be drawn by linking the first maxima in each sentence on the one hand, and the first minima in each sentence on the other.

However, last maxima and minima for sentences seem quite similar through the paragraphs. This hypothesis, tested already for English (Ladd [5]) and Danish (Grønnum Thorsen [3,4]), seems to be also valid for Spanish.

The analysis of local phenomena shows that Fo maxima do not appear to be an important cue to mark stressed syllables in reading style: about 80 % of the stressed syllables does not match with Fo maxima. It is worth noting here that maxima appear primarily in the syllable following the stressed one in all conditions.

Duration could be an important cue for lexical stress in reading style in Spanish (see also Canellada & Kuhlmann [6]). The differences in duration found between unstressed and stressed syllables seem to support this idea.

Parameters related to the global shape of the Fo contour -specially to its initial part- together with the range seem to be the most important in order to discriminate between the conditions studied. The topline seems also important for discriminating among conditions. Inversely, parameters related to the bottomline and to the final part of the contour appear to have more invariant values.

It seems that the parameters related to the global shape of the contour can be arranged in order of importance according to the number of conditions they are able to discriminate:  $P_I$ , the difference between  $P_I$  and  $V_I$  and the range are the most important parameters. The slope of the topline and the frequency of  $V_I$  appear to be the second most important ones. The third most important parameters are  $P_F$  and the slope of the bottomline.

When local parameters are studied, it is observed that the distribution of Fo maxima and minima related to stress and to the duration of syllables is not significantly different in the four conditions examined.

Although no significant differences are found between isolated and text initial conditions, some differences between

isolated on the one hand and text medial and text final sentences on the other have been observed. These differences are found in the top declination line ( $P_I$  and  $P_F$ , slope, difference between  $P_I$  and  $V_I$ ) and in the range, but are not observed in the parameters related to stress.

These results agree with those reported by Thorsen [3], who obtained similar values for isolated and text initial sentences. Lehiste [2] states that Fo values at the beginning of isolated sentences are lower than in paragraph initial ones; the findings in this paper show the same tendency, although they appear as statistically not significant.

Differences similar to those existing between isolated and contextualized sentences have been found among sentences in different positions in the paragraph:  $P_I$ , the difference between  $P_I$  and  $V_I$ , and the range seem the best cues to distinguish among these conditions. Lehiste [2], Thorsen [3, 4] and Ladd [5] report analogous results.

## REFERENCES

- [1]: Umeda, N. (1982) "F0 declination is situation dependent", *Journal of Phonetics* 10: 279-290.
- [2]: Lehiste, I. (1984) "The phonetic structure of paragraphs" in Cohen, A.- Nooteboom, S.G. (Eds.) *Structure and Processes in Speech Perception*. Berlin: Springer. pp. 195-206.
- [3]: Grønnum Thorsen, N. (1985) "Intonation and text in Standard Danish", *Journal of the Acoustical Society of America* 77, 3: 1205-1216.
- [4]: Grønnum Thorsen, N. (1986) "Sentence intonation in textual context - supplementary data", *Journal of the Acoustical Society of America* 80,4: 1041-1047.
- [5]: Ladd, D.R. (1988) "Declination 'reset' and the hierarchical organization of utterances", *Journal of the Acoustical Society of America* 84,2:530-544.
- [6]: Canellada, M.J.- Kuhlmann, J. (1987) *Pronunciación del español. Lengua hablada y literaria*. Madrid: Castalia.

## ACKNOWLEDGMENT

This work has been supported by a grant from the *Dirección General de Investigación Científica y Técnica* (PB90-0704)

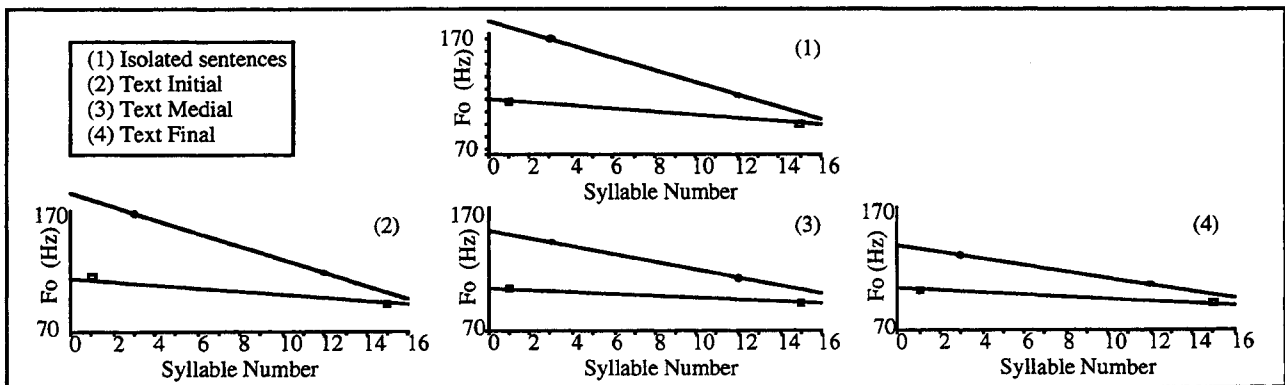


Figure 1:  
Top and bottom lines calculated using  $P_I$  and  $P_F$  (topline) and  $V_I$  and  $V_F$  (bottomline) averaged values