Labeling Melodic Movements at the Stress Group Level*

Lourdes Aguilar
David Casacuberta
Rafael Marín
Universitat Autònoma de Barcelona. Departament de Filologia Espanyola
08193 Bellaterra (Barcelona). Spain
Lourdes.Aguilar@uab.es
dcasacuberta@seneca.uab.es
rmarin@sumi.es

Received: July 15th 1998
Accepted: January 10th 1999

Abstract

The aim of this work is to propose a system of F0 labeling using the stress group as the prosodic unit that can accurately relate linguistic and acoustic information. To obtain data, a corpus of sentences is given to read to a male speaker and the melodic information contained in the vowels is analyzed. The basic melodic movements —encoded as rise, fall and connection elements— between stress groups in an intonation group are observed in order to develop a set of rules assigning them from text.

Results show that by means of the stress group it is possible to model some prosodic facts, such as the effect of sentence length on the height of the first F0 peak in the sentence, as well as to incorporate the syntactic information encoded in the lexical labeling of the stress group.

Key words: intonation, stress group, tonal assignment, melodic movements.

Resum. *L'etiquetatge de moviments melòdics a partir del grup accentual*

En aquest treball proposem un sistema d'etiquetatge de F0 que parteix del grup accentual com a unitat prosòdica que posa en relació la informació lingüística i l'acústica. Per obtenir dades, partim d'un corpus de frases llegides per un parlant natiu i analitzem la informació melòdica de les vocals. Observem els moviments melòdics bàsics —codificats en forma de moviments d'ascens, descens o continuació tonal— entre grups accentuals en un grup entonatiu per tal de desenvolupar un sistema de regles que puguin assignar aquests moviments a partir d'un text.

Els resultats demostren que és possible, mitjançant el grup accentual, d'obtenir una descripció adequada de la prosòdia, que inclou l'efecte de llargada de frase en l'altura del primer pic F0 de la frase, així com d'incorporar la informació sintàctica codificada en l'etiquetatge lèxic del grup accentual.

Paraules clau: entonació, grup accentual, assignació tonal, moviments melòdics.

[*] Part of this work has been presented at the *LabPhon V. Fifth Conference on Laboratory Phonology*, at Evanston, USA, July 6th-8th 1996 and at the *XXVII Simposio de la Sociedad Española de Lingüística* held at Palma de Mallorca, December 15th-19th 1997. R. Marín's work has been partially supported by a grant from Generalitat de Catalunya (AP-LP/95-2704). We wish to thank the anonymous reviewers for comments and criticism.

Table of Contents

1. Introduction

When dealing with speech technology problems, it is often asserted that the global quality of a text-to-speech system can be improved with an adequate prosodic modeling. However, the content and the extent of this modeling is not commonly acknowledged. Moreover, the undeniable relations of prosody with other disciplines such as phonology, syntax, semantics or pragmatics add difficulties to the treatment of this subject.

We could mention two core problems in the field of intonational theory: firstly, the problem of delimiting the linguistic information needed for modeling prosody, and secondly, how to relate this linguistic information to the acoustic one.

With regard to the first question, several approaches concerning the type and extent of syntactic information are found: from positions arguing in favor of an identity of syntactic and prosodic structures (Liberman and Prince, 1977; Cooper and Paccia-Cooper, 1980; Umeda, 1982) to those theories that propose a relationship between prosody and syntax by means of an intermediate level (O'Shaughnessy, 1990; Bachenko and Fitzpatrick, 1990; Hirschberg and Prieto, 1996; Ladd, 1996). In addition to this, a strong link with semantic and pragmatic information is being highlighted recently (Prevost, 1995; Veilleux, 1997; Hiyakumoto *et al.*, 1997).

Nevertheless, most theories treat the prosodic and syntactic structures of utterances as related but not identical. For example, Selkirk (1984) argued for the prosodic structure as a distinct entity with a defined relationship to syntactic structure. She built a prosodic structure by grouping words into phonological words, and then into larger prosodic units using syntactic information, providing then prosodic structures which were not necessarily isomorphic to syntactic constituents. Phonetic evidence has been adduced by authors such as Thorsen (1981): in syntactically unambiguous non-compound sentences, the prosodic stress group cuts accross any syntactic boundary, and in long utterances, the prosodic phrasing bears no simple relation to syntactic structure.

Concerning the problem of relating linguistic and acoustic information, most of the existing models provide either a non-formal or an incomplete account of the interaction between these levels (see Taylor, 1992, for a detailed review). Before proposing a phonetic model of intonation for English, Taylor (1992) reviews some of the existing phonetic models; namely, that of the British school (Halliday, 1967; Crystal, 1969), the Dutch school (t'Hart & Cohen, 1973; t'Hart & Collier, 1975), the Pierrehumbert model (Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988) and the Fujisaki model (Fujisaki & Kawai, 1982). It is shown that these models are not entirely satisfactory, either because of the difficulty to provide an F0-phonology mapping —as in the case of the British school or the Pierrehumbert approach— or because they are not sufficiently powerful to model all the occurring types of F0 contours, which is the case for the Dutch school or the Fujisaki model.

To tackle the phonetic modeling task, Taylor (1992) uses a formal approach, which assumes description of systems in terms of levels, grammars and mappings. This model uses three levels being linked by means of grammars: an F0 level, an intermediate level based on the principle that F0 contours can be described as a sequence of rise, fall and connection elements, and a phonological level.

Although being inspired by all the preceding work in the field, the aim of this study is somewhat less ambicious. What we are mainly concerned with is a phonologically-oriented description suitable for generating proper sentence prosody in Spanish, by means of the association of some simple melodic movements to linguistic units.

To do this, we will rely on the text analysis procedure used in previous studies on pause assignment (Marín *et al.*, 1996; Casacuberta *et al.*, 1997). According to it, a text can be segmented using a syntactico-prosodic unit which serves to write a set of rules to assign pauses. We call this unit categorial stress group (CSG) and can be defined as a syntactically labeled stress group (SG) —a stressed word preceded, if appearing, by one or more unstressed words—[1].

But since not only pauses segment the speech chain prosodically, it is necessary as well to consider the F0 movements where syntactic information may be involved. According to previous work in the field, the combination of pauses and F0 movements should divide a sequence into major and minor prosodic groups (Cruttenden, 1986; Wang and Hirscherg, 1992; Steedman, 1991).

Our main interest is to use the CSG for proposing a set of rules that can explain both the pause assignment and the main F0 movements in intonation groups (a stretch of speech delimited by pauses). To do this, we assume that F0 assignment can be modeled by means of the SG, which has been described as a phonological domain, where some segmental and prosodic processes take place (Nespor and Vogel, 1986; Levelt, 1989). Thus, it is intended to develop a battery of rules concerning the F0 movements in an intonation group using the grammatical information associated to the SG, which is encoded in the CSG unit. These rules are derived from the results of an experimental analysis. A corpus of sentences is given to read to a male speaker and the melodic information contained in the vowels is analyzed.

The basic melodic movements between SGs belonging to an intonation group are observed in order to develop a set of rules assigning them from text, and the F0 values associated to sentence positions are analyzed so as to provide a quantitative framework in which F0 variations apply.

[1] Note that our definition is different from others such as the proposed by Thorsen (1991): «a prosodic stress group is defined as a stressed syllable plus all succeeding unstressed syllables, irrespective of intervening syntactic boundaries within the same intonation contour».

2. Procedure

Before analyzing the melodic movements associated to SGs, it is necessary to describe the text analysis procedure used to map the acoustic and the linguistic level.

2.1. Text labeling

Every word of the text belongs to a lexical category according to the following list: adjective (a), adverb (ad), conjunction (c), coordinating conjunction (cc), clitic (cl), gerund (g), infinitive (i), noun (n), preposition (p), participle (pt), quantifier (q)[2], verb (v). Each category includes information referring to lexical stress which relies on the distinction between open and closed categories: open categories (a, ad, g, i, n, pt, v) are considered to be stressed whereas closed ones (c, cc, cl, p, q) are not.

Later, the string of categories are grouped in order to conform CSGs, that is, SGs which are labeled according to the lexical category of the first element in the group[3].

Table I presents the list of CSGs with their associated syntactic head. It must be said that some CSGs are formed by just one element (for instance, vg, ng or ag) whereas others have more than one element (clg, ccg or qg): as for the latter, we will refer to the first element as head and to the last one, as modifier.

Table I. List of CSGs and their syntactic head.

CSG	syntactic head
ag	adjective
adg	adverb
cg	conjunction
ccg	coordinating conjunction
clg	clitic
gg	gerund
ig	infinitive
ng	noun
pg	preposition
ptg	participle
qg	quantifier
vg	verb

[2] We use ‘quantifier’ in a broad sense, i.e., including categories such as articles, demonstratives or possessives.

[3] The use of this unit bears certain resemblance to the idea of parsing by chunks proposed in Abney (1992).

To illustrate the text segmentation procedure, a CSG labeling is offered in (1), where modifiers appear between brackets[4].

(1) [Las mujeres]qg(n) [de las casas]pg(n) [inundadas]ag
[the women]qg(n) [from the houses]pg(n) [flooded]ag
[encontraron]vg [el camino]qg(n)
[found]vg [the way]qg(n)
'The women from the flooded houses found the way.'

2.2. Aim

In order to obtain information about the interaction between melodic movements and the linguistic information encoded in CSGs, a corpus of sentences has been analyzed. Our main aim is to ellucidate if the syntactic properties of a given CSG triggers a change in the F0 shape of those surrounding it. Related to this, some questions can be raised.

On the one hand, we want to find out if the main F0 movements syntactically conditioned appear within CSGs or, on the contrary, they occur at the boundaries between two CSGs. To determine this, F0 trajectories from the last syllable of a CSG to the first syllable of the following CSG have been observed. It is hypothesized that CSGs behave as a syntactico-prosodic domain, in which main prosodic changes are manifested; this implies that great differences between F0 movements at CSG boundaries will not be expected.

On the other hand, we want to know if there are contextual effects in a given string of CSGs: more specifically, if a given CSG triggers a change in the F0 contour of the preceding CSG in the sentence or, on the contrary, in the following one. To elucidate this, we adopt the following procedure: taking as landmark a certain type of CSG, the melodic movement just before and after it is observed by analyzing the F0 movements in the final stage of the preceding CSG and in the initial stage of the following CSG. For the sake of homogenizing CSG syllabic lengths, we consider the two last syllables of the CSG as its final stage and the two first syllables as its initial stage.

2.3. Corpus

A set of 140 declarative sentences, with SVO order and no subordinate clause inside them, constitutes the core of the analysis. Sentences have been gathered from a large corpus of scientific and literary texts and paragraphs including the sentences have been rewritten in order to conform three texts.

[4] See Casacuberta *et al.* (1997, 1998) for a more detailed description.

2.4. Analysis

Texts were read by a Spanish speaker (who comes from the central area of Spain), and recorded in a digital tape. Recordings took place in a soundproof room using a Sony DTC790 digital tape recorder and a Sennheiser MKH20 microphone. The speaker did not receive any instruction about his reading but he used his own speech rate and range of tones.

For all the sentences of the corpus, F0 contours were obtained using the MacSpeechLab II software, with a pitch tracking algorithm based on an autocorrelation technique. The center of the vowel was determined from the waveform display and the F0 value measured. A total number of 3245 vowels was analyzed.

The CSG transcription was aligned with the recorded speech signal, and for each type of CSG, the following data were calculated: first, the difference between the F0 values of the two vowels situated at the CSG boundaries; second, the difference between the F0 values of the two last vowels in the preceding CSG; and finally, the difference between the two F0 values of the two first vowels in the following CSG. Adopting this procedure, in which contextual effects are being observed, we are obliged to analyze only CSGs appearing in a medial-phrase position.

A total number of 848 items (424 CSGs x 2 F0 vowel values) has been computed when facing with the preceding CSG, and a number of 776 items (388 CSGs x 2 F0 vowel values) when analyzing the following CSG. Results are presented in absolute and relative values, and statistical tests are applied on data to determine the significance of the differences.

Following Taylor (1992), we assume that F0 contours can be described as a sequence of rise, fall and connection elements. Thus, from a minimal difference of 5 Hz., three movements have been established: rise (R), if the difference is positive, fall (F), if the difference is negative, and connection (C), if there is no difference[5].

3. Results

3.1. Melodic movements at CSG boundaries

Table II presents the number of occurrences of each type of CSG[6] and the F0 movement associated to a given CSG boundary: R, S or C. As said before, these results were obtained calculating the difference between the F0 value of the last vowel of a CSG and the F0 value of the first vowel of the CSG that followed it. Groups are sorted according to the frequence of occurrence of F element.

[5] Our labeling is similar to the one used by Taylor (1992), although the analysis is different. This author uses fall shapes to describe the falling parts of pitch accents; rise shapes to describe both the rising parts of pitch accents and rises observed at the beginnings and ends of phrases; finally, connection elements are used everywhere else.

[6] Neither ccg nor gg are included because there were not enough cases available in the corpus.

Table II. Number of occurrences of each type of CSG associated to F0 movements at CSG boundaries.

		Rise (R)		Fall (F)		Connection (C)	
	Total n	**n**	**%**	**n**	**%**	**n**	**%**
cg	10	—	—	9	90	1	10
clg	13	—	—	11	85	2	15
vg	45	—	—	33	73	12	27
ptg	27	1	4	20	71	6	25
adg	26	—	—	16	62	10	38
ag	56	3	5	34	61	19	34
pg	149	9	6	88	59	52	35
ig	16	—	—	9	56	7	44
ng	28	4	14	14	50	10	36
qg	54	5	9	19	35	30	56

It can be observed that there is an important tendency to the falling movement. From a total number of 424 items, only in 22 CSG boundaries an R element (5%) has been found; on the contrary, in 253 boundaries, an F element is identified (60%), the rest corresponding to the C one (35%).

Differences are significant, as inferred from a chi-square test applied on the data (chi-square: 37.4, $p < 0.05$).

3.2. Melodic movements within the preceding CSG

Table III presents the total number of the analyzed CSGs, and, for each type of CSG, the number of cases of R, F and C in the preceding CSG. These results were obtained calculating the difference in the F0 values of the two last vowels of the CSG that precedes the analyzed CSG.

From a total of 424 CSGs, 162 are associated to an R element (38%), 101 to an F element (24%) and 161 to a C element (38%). Differences between groups are significant, as inferred from a chi-square test applied on the data (chi-square: 43, $p < 0.05$). Groups are sorted according to the frequency of occurrence of R element.

As can be seen in Table III, the typical melodic movement found before cg, clg, ng, pg and vg is R while ptg and qg show a high percentage of connection elements before them. On the other hand, there is not a strong preference towards one of the melodic movements before adg, ag and ig.

Although being aware that these results are preliminary, it can be said that a given CSG, according to its syntactic properties, affects prosodically the SG which precedes it. It could be hypothesized that before a group with a strong semantic load, as in which CSGs where the stressed word is a verb or a noun —which is the case for cg, clg, ng, pg and vg— a rise movement appears in order to identify information structure. Conversely, other groups such as ag, adg or ig, typically linked to

Table III. Number of occurrences of each type of CSG associated to F0 movements within the preceding CSG.

		Rise (R)		Fall (F)		Connection (C)	
	Total n	n	%	n	%	n	%
cg	10	7	70	1	10	2	20
clg	13	8	62	1	8	4	31
vg	45	25	56	6	13	14	31
ig	16	8	50	—	—	8	50
ng	28	13	46	6	21	9	32
pg	149	62	42	36	24	51	34
adg	26	10	38	6	23	10	38
qg	54	15	28	13	24	26	48
ptg	27	5	18	9	33	13	48
ag	56	9	16	23	41	24	43

other syntactic groups for transfering content, are subject to the general declination effect.

3.3. *Melodic movements within the following CSG*

Table IV presents for each type of CSG the total number of CSGs and the number of cases showing an F0 movement of R, F or C in the initial phase of the following CSG. As said before, these labels have been determined by means of the differ-

Table IV. Number of occurrences of each type of CSG associated to F0 movements within the following CSG.

		Rise (R)		Fall (F)		Connection (C)	
	Total n	n	%	n	%	n	%
ptg	20	3	15	15	75	2	10
ng	26	4	15	17	65	5	19
qg	54	9	17	32	59	13	25
vg	62	10	16	36	58	16	26
pg	116	18	15	64	55	34	29
ig	13	—	—	7	54	6	46
clg	19	1	5	10	53	8	42
cg	14	1	7	7	50	6	43
adg	35	2	6	17	49	16	46
ag	29	6	21	9	31	14	48

ence observed between the two first vowels of this CSG. Groups are sorted according to the frequency of occurrence of F element.

From a total number of 388 CSGs, 214 have been associated to F (55%), 120 have not shown changes in their F0 trajectory (31%) and only in 54 cases a rising F0 has been found (14%). Differences between groups are not significant, as inferred from a chi-square test applied on data (chi-square: 25.4, $p > 0.05$).

3.4. Assigning F0 labels

From the information obtained in the experimental analysis, a simple set of rules can be derived to prosodically label a text. Two main trends have arisen. First, an F movement appears both between two consecutive CSGs and at the beginning of a CSG. And second, syntactic information seems to be relevant only when assigning F0 contours to the final part of a given CSG: the appearance of F, R or C elements depends on the linguistic information encoded in the CSG. For instance, a falling movement is associated to the last syllables of a CSG preceding an ag while a rising movement is assigned to the last syllables of a CSG preceding a vg.

As a consequence of the interaction of syntactic and prosodic factors, we can develop a simple model of prosodic labeling. At this stage of the work, the following rules can be proposed:

1) Assign an F label between CSG boundaries.
2) If the CSG is phrase-medial, assign an F label to its initial part.
3) Look at the type of the following CSG and assign the corresponding F0 label to the final part of the CSG:
   3a) assign an R label if the CSG is followed by ng, pg, vg, clg, and cg;
   3b) assign a C label if the CSG is followed by qg or ptg;
   3c) assign at random R, F or C if the CSG is followed by adg or ig;
   3d) assign at random F or C if the CSG is followed by ag.

4. Anchoring F0 values

So far we have discussed how labels such as rising F0 movement, falling F0 movement or continuation F0 movement can be related to strings of text. However, the reinterpretation in quantitative values is what applications such as text-to-speech systems are mainly concerned with. The following experiment is designed to ascertain if it is possible to find some positions in the sentence that serve as the reference points from which the R, F and C labels could be interpreted. Since length constraints are well known in intonation, we will try to examine the relationship between three points of the F0 contour and the number of SGs contained in the sentence. The selected points are: the initial F0 value, considered to be the F0 value of the first vowel in the sentence, the first F0 peak, which is the F0 maximum in the sentence, and the final F0 value, corresponding to the F0 value of the last vowel.

We are aware that it is customary to use the number of syllables contained in the sentences in order to study the effect of sentence length. However, if we want to be consistent with the use of the SG to model intonation, it is necessary to see if measuring sentence length in SGs is a good predictor index. Thus, we present results according both to the number of syllables and to the number of SGs.

As for the first, we follow traditional Spanish intonational descriptions (Navarro Tomás, 1948; Canellada & Kuhlmann, 1987) to distinguish three categories of sentences depending on the number of syllables contained in them: *a)* short sentences, with 5 or less than 5 syllables; *b)* medium sentences, with more than 5 and less than 13 syllables; *c)* long sentences, with more than 13 syllables. In comparison, depending on the number of SGs contained in sentences, we differentiate three groups: *a)* short, with two or less than two SGs; *b)* medium, with more than two and less than six SGs; *c)* long, with six or more than six SGs.

Results in function of these criteria can be compared in Tables V and VI, in which the number of cases (n), the mean values (x) and standard deviation (sd) of F0 values corresponding to the initial point, the F0 maximum and the final point of the sentence are presented.

Data depend on three criteria. First, if sentences initially classified as short, medium or long have a pause inside them, the sentence is discarded. Second, if the initial F0 value coincides with the first peak, values are duplicated. And third, if there is final-devoicing, no value for final F0 is taken.

It can be observed that independently of the unit of measure (syllable or SG), initial F0 values and first F0 peak values show higher values when sentence length is increased. However, only differences concerning the F0 maximum value are significant. An ANOVA applied on data grouped according to the number of syllables of the sentence does not find differences either in initial F0 values (F = 3.5, p > 0.01) or in final F0 values (F = 3.2, p > 0.01). In contrast, the same analysis

Table V. Effect of sentence length (measured in number of syllables) on initial, maximum and final F0 values.

	Initial F0 value			**First F0 peak value**			**Final F0 value**		
	n	**x**	**sd**	**n**	**x**	**sd**	**n**	**x**	**sd**
short sentence	20	118	8	20	123	10	13	108	7
medium sentence	40	118	10	40	140	10	17	104	6
long sentence	43	124	13	43	145	22	29	103	5

Table VI. Effect of sentence length (measured in number of SGs) on initial, maximum and final F0 values.

	Initial F0 value			**First F0 peak value**			**Final F0 value**		
	n	**x**	**sd**	**n**	**x**	**sd**	**n**	**x**	**sd**
short sentence	22	118	8	22	125	11	13	106	5
medium sentence	50	120	11	50	142	11	25	105	8
long sentence	31	124	13	31	143	25	21	102	5

points out significant differences between the groups (F = 12.2, p < 0.01) in data referred to the first F0 maximum in the sentence. Nevertheless, it has to be pointed out that a Scheffe-test identifies differences in the pairs short *vs*. medium (6.8) and short *vs*. long (12.1) but not in the pair medium *vs*. long

A similar behavior is found with data grouped according to the number of SGs: the differences arise in F0 peak values and not at the beginning or the end of the sentence. A one-way ANOVA does not reveal significant differences either in initial (F= 1.98, p > 0.01) or final F0 values (F= 1.6, p > 0.01) whereas significant differences between the groups arise in data concerning to the F0 maximum value (F = 9, p < 0.01). In this case, however, as we have also reported for lenght measured in syllables, a Scheffe-test only points out differences between the pairs short vs. medium (7.5) and short vs. long (7.2).

It can be concluded that while both initial and final F0 value remain constant, F0 maximum changes so as to be adapted to sentence length, measured either in number of syllables or SGs. Moreover, the results suggest that sentence length distinguishes between two groups —short and long sentences, the later including medium sentences— instead of three, as proposed in Navarro Tomás (1948) or Canellada & Kuhlmann (1987).

5. An example of F0 assignment

From the results in previous sections, a battery of rules and some reference F0 points are used to develop an intonational grammar. Now, we offer an illustration of how it works.

Once the text has been segmented into CSGs, the rules proposed in 3.4 are applied in order to model factors such as length constraints or contextual influences between SGs.

In order to incorporate this information into a simple model, three points are used as reference: *a)* the initial F0 value; *b)* the first peak in the sentence; and *c)* the final F0 value. Initial and final F0 value are linked respectively to the first and last vowels in the sentence, whereas the first F0 maximum in the sentence is associated to the vowel of the last syllable in the first CSG its value being strongly dependent on sentence length. According to this, it is worth pointing out that in the corpus, the first F0 maximum in the sentence tends to coincide with the last vowel of the first SG (75%), thus supporting the segmentation into SGs.

Let us see how the model works on a sentence like *Las mujeres de las casas inundadas encontraron el camino* (‘The women from the flooded houses found the way’). The first step is to parse the sentence into CSGs as shown in (2):

(2) [Las mujeres]qg(n) [de las casas]pg(n) [inundadas]ag [encontraron]vg
[el camino]qg(n)

Since the sentence has five SGs, it is considered to be medium and, therefore, the corresponding values are applied to obtain initial, final and first F0 peak values.

Following the observations concerning the location of the first F0 maximum, the model associates, as a first approximation, this peak to the vowel of the last syllable in the first CSG.

Next, the rules related to F0 assignment are applied:

1. Assign an F label to the boundary of each pair of CSGs; that is, the F0 value of the first syllable of a given CSG is lower than the F0 value of the last syllable of the preceding CSG:

```
                    F           F          F            F
                  • -----•    • ---•     • ---•       • ----•
(3)   [Las mujeres] [de las casas] [inundadas] [encontraron] [el camino]
```

2. Assign an F label to the initial part of each CSG; in other words, the F0 value of the second syllable of a given CSG has to be lower than the F0 value of the first syllable of this CSG:

```
                       F          F          F
                     • -•       • -•       •---•
(4)   [Las mujeres] [de las casas] [inundadas] [encontraron] [el camino]
```

3. Look at the type of CSG and assign the corresponding F0 label to the final part of the preceding CSG. In this case, since *inundadas* is a ptg, the last two syllables of *de las casas* have a connection movement. Likewise, since *encontraron* is a vg, the last two syllables of *inundadas* have a rise movement:

```
                               C          R           C
                             • -•       • -•        • -•
(5)   [Las mujeres] [de las casas] [inundadas] [encontraron] [el camino]
```

The final result of the application of these rules is as follows:

```
                    F   F     C  F  F    R  F  F      C  F
                  • -----• -•  • -• ---• -•  •--• ---•---•   • -• ----•
(6)   [Las mujeres] [de las casas] [inundadas] [encontraron] [el camino]
```

6. Conclusions

The main results of the preceding experiments can be summarized as follows.

— First, neither initial nor final F0 values depend on sentence length.

— Second, the effect of sentence length is manifested on the height of the F0 maximum in the sentence, which tends to coincide with the last vowel of the first CSG.

— Third, the more relevant F0 movements appear within CSGs, not at CSG boundaries. It is inferred from this that melodic movements have to coincide with a CSG boundary.

— Fourth, F0 movements appearing at the final part of a CSG are chiefly determined by the syntactic information conveyed by the CSG that follows it.

From the data, prosodic and syntactic factors affecting F0 assignment can be described. Prosodic factors are mainly concerned with the height of the first F0 maximum in the sentence and due to length constraints, whereas syntactic factors are manifested in the medial part of the sentence by means of the influence of the type of CSG in F0 contours.

The results of the experiment can be easily coded in an experimental protocol merging information about pause assignment and melodic movements. The needed information would be:

a) the modality of each sentence to be uttered (declarative, interrogative, exclamative);[7]
b) a list of all words to be uttered along with their associated categories;
c) an specification of each group along with its associated group category;
d) a description of the connections between groups that will inform about:
   d.1) the position of pauses;
   d.2) the melodic movement within intonation groups.

Despite various shortcomings, such as the need to enlarge the data bulk, or to include more syntactic variables, we have demonstrated in this study that F0 contours can be described using the SG as a basic unit. By means of this unit, it is possible to observe and model relevant prosodic facts: the effect of sentence length on the height of the first F0 peak in the sentence is a case in point. It is also feasible to describe a set of melodic movements depending on the linguistic information provided by the CSG previous to a given CSG. Related to this, it could be said that F0 movements are domain-governed and contextually conditioned.

Once again, as shown in Casacuberta *et al.* (in press) for pause location, detailed information concerning both prosodic and syntactic factors is needed. Therefore, by using CSGs it is possible to predict not only pauses, but also melodic movements. A syntactic labeling of prosodic units allows us to propose a model that relates in a direct way melodic movements and linguistic units. This model represents a simplification over other existing models that handle syntactic and prosodic aspects of speech separately, making its implementation in a text-to-speech system a feasible fact. Nevertheless, the great number of connection movements found in the corpus analyzed suggests that CSGs should be combined to obtain a higher-level unit in which F0 contours are organized. Moreover, an effort on quantitative description has to be expended before incorporating the model in a text-to-speech system.

[7] At this moment, only information concerning declarative sentences is being offered, but it should be complemented with studies on other modalities.

References

Bachenko, J.; Fitzpatrick, E. (1990). «A computational grammar of discourse-neutral prosodic phrasing in English». *Computational Linguistics,* 16(3): 155-170.

Canellada, M.J.; Kuhlmann, J. (1987). *Pronunciación del español.* Madrid: Castalia.

Casacuberta, D.; Aguilar, L.; Marín, R. (1997). «ProPause: a syntactico-prosodic system designed to assign pauses». *Proceedings of Eurospeech'97*. Rhodes, Greece, 1: 203-206.

Casacuberta, D.; Marín, R.; Aguilar, L. (1998). «Parsing unrestricted text into prosodic units». In C. Martín Vide (ed.). *Mathematical and Computational Analysis of Natural Language. Studies in Functional and Structural Linguistics.* Amsterdam: John Benjamins, 45: 281-294.

Cooper, W.E.; Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge: Harvard University Press.

Cruttenden, A. (1986). *Intonation.* Cambridge: Cambridge University Press.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Fujisaki, H.; Kawai, H. (1988). «Realization of linguistic information in the voice fundamental frequency contour of the spoken Japanese». In *International Conference on Speech and Signal Processing,* IEEE.

Halliday, M.A.K. (1967). *Intonation and Grammar in British English*. The Hague: Mouton.

Hirschberg, J.; Prieto, P. (1996). «Training intonational phrasing rules automatically for English and Spanish text-to-speech». *Speech Communication,* 18(3): 283-290.

Hiyakumoto, L.; Prevost, S.; Cassell, J. (1997). «Semantic and discourse information for text-to-speech intonation». *Proceedings of the Concept to Speech Generation Systems.* Madrid.

Ladd, R. (1996). *Intonational Phonology.* Cambridge: Cambridge University Press.

Levelt, W.J.M. (1989). *Speaking: from Intention to Articulation*. Cambridge: The MIT Press.

Liberman, M.Y.; Prince, A. (1977). «On stress and linguistic rhythm». *Linguistic Inquiry*, 8: 249-336.

Marín, R.; Aguilar, L.; Casacuberta, D. (1996). «El grupo acentual categorizado como unidad de análisis sintáctico-prosódico». In C. Martín Vide (ed.), *Lenguajes Naturales y Lenguajes Formales,* XII. Barcelona: PPU, 487-494.

Navarro Tomás, T. (1948). *Manual de entonación española.* Madrid: Guadarrama.

Nespor, M.; Vogel, I. (1986). *Prosodic Phonology.* Dordrecht: Foris.

O'Shaughnessy, D. (1990). «Relations between syntax and prosody for speech synthesis». *Proceedings of the ESCA Workshop on Speech Synthesis.* Autrans, France.

Pierrehumbert, J.B. (1980). *The Phonology and Phonetics of English Intonation*. Bloominghton: Indiana University Linguistics Club.

Pierrehumbert, J.B.; Beckman, M. (1988). *Japanese Tone Structure.* Cambridge: The MIT Press.

Pierrehumbert, P. (1980). *The Phonology and Phonetics of English Intonation.* MIT. Doctoral Dissertation.

Prevost, S. (1995). *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation.* University of Pennsylvania. Doctoral Dissertation.

Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge: The MIT Press.

Steedman, M. (1991). «Structure and intonation». *Language*, 67(2): 260-296.

Taylor, P. (1992). *A Phonetic Model of English Intonation*. Edinburgh. Doctoral Dissertation.

t'Hart, J.; Cohen, A. (1973). «Intonation by rule: a perceptual quest». *Journal of Phonetics*, 1: 309-327.

t'Hart, J.; Collier, R. (1975). «Integrating different levels of intonation analysis». *Journal of Phonetic*s, 3: 235-255.

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.

Veilleux, N. (1997). «Probabilistic model of acoustic / prosody / concept relationships for speech synthesis». *Proceedings of the Concept to Speech Generation Systems.* Madrid, 1-10.

Wang, M.Q.; Hirschberg, J. (1992). «Automatic classification of intonational phrase boundaries». *Computer Speech and Language*, 6: 175-196.